(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number
WO 01/55450 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/02474

(22) International Filing Date: 24 January 2001 (24.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/179,061 28 January 2000 (28.01.2000) US

(71) Applicant *(for all designated States except US)*: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES, THE NATIONAL INSTITUTES OF HEALTH** [US/US]; Office of Technology Transfer, Suite #325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KALLIONIEMI, Anne, H.** [FI/US]; 1083 Grand Oak Way, Rockville, MD 20852 (US). **MONNI, Outi, M.** [FI/US]; 6020 California Circle #414, Rockville, MD 20852 (US). **BÄRLUND, Maarit, T.** [FI/FI]; Ilmarinkatu 19 B19, FIN-33500 Tampere (FI). **KONONEN, Juha, T.** [FI/US]; 1920 Valley Stream Drive, Rockville, MD 20851 (US). **KALLIONIEMI, Olli, P.** [FI/US]; 1083 Grand Oak Way, Rockville, MD 20852 (US).

(74) Agent: **HARDING, Tanya, M.**; Klarquist, Sparkman, Campbell, Leigh & Whinston, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/55450 A2

(54) Title: AMPLIFIED GENES AT 17q23

(57) Abstract: The present disclosure provides new highly sensitive probes for the detection of neoplasms, including breast and other cancers. The probes bind selectively with target polynucleotide sequences found on the long arm of Chromosome 17, more particularly the region 17q23. In particular, such probes are to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449. Amplification or other mutation of any of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 is frequently linked to cancer. Methods of using RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 biological materials in the diagnosis and treatment of neoplasia, and particularly breast cancer, are presented.

AMPLIFIED GENES AT 17q23

FIELD

This disclosure relates to the field of human genetics. More particularly, the present disclosure relates to the linkage of certain genes within the 17q23 chromosomal region to human cancer. Genomic amplification of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 is frequently linked to cancer, particularly breast cancer. The present disclosure also relates to methods for screening and diagnosis of neoplasia, *e.g.* breast cancer, and methods of gene therapy utilizing recombinant DNA technologies. It also provides new diagnostic nucleic acid markers for breast and other cancer.

BACKGROUND

The isolation of genes involved in human cancer development is critical for uncovering the molecular basis of cancer, in which unregulated or improperly regulated clonal proliferation of a cell population occurs. Gene amplification plays an important role in the progression and initiation of many solid tumors, including breast cancer. In gene amplification, the copy number of a genomic DNA sequence is increased in a cancer cell compared to a non-cancerous cell. To date, at least 20 genes, including HER-2, CCND1, EMS1, MYC, EGFR, FGFR1, FGFR2, and AIB1, are amplified in breast cancer. Some of these amplifications, such as HER-2 at 17q12-q21, as well as CCND1 at 11q13, and AIB1 at 20q12, have been linked to poor prognosis of the patients.

Recently, molecular cytogenetic studies have revealed the occurrence in breast cancers of additional regions of increased DNA copy number (Isola *et al., Am. J. Pathol.* 147:905-911, 1995; Kallioniemi *et al., Proc. Natl. Acad. Sci. USA* 91:2156-2160, 1994; Muleris *et al., Genes Chromo. Cancer* 10:160-170, 1994; Tanner *et al., Cancer Research* 54:4257-4260, 1994; Guan *et al., Nat. Genet.* 8:155-161, 1994). One commonly amplified region, 17q22-q24, has been shown to be amplified in about 20% of primary breast tumors by CGH (Courjal and Theillet, *Cancer Res.*, 57:4368-4377, 1997; Tirkkonen *et al., Genes Chromo. Cancer*, 21:177-184, 1998). Recently, research has indicated that the 17q22-q24 amplification in breast cancer is due to high-level amplification of at least two separate regions. One of these amplified regions has been more distinctly localized to 17q23 (Bärlund *et al., Genes Chromo. Cancer*, 20:372-376, 1997; Pollack *et al., Nat. Genet.*, 23:41-46, 1999). The 17q23 region is relatively gene rich.

Other genes located within the 17q23 region have been identified because of their native function. These include, for instance, ribosomal protein S6 kinase (S6K), a critical mediator involved in G1 to S-phase progression; RAD51C, a strand-transfer protein hypothesized to be involved in meiotic recombination; PAT1 (also known as APPBP2), a cytoplasmic protein implicated in translocation of amyloid precursor protein along microtubules; MUL, a member of the RING-B-

box-Coiled-coil family of proteins implicated in the genetic recessive disease mulibrey nanism; TRAP240, a coactivator to the nuclear hormone receptor family; and TBX2, a T-box transcription factor gene. No research has linked any of the RAD51C, PAT1, MUL, TRAP240, or TBX2 genes to breast cancer, nor has genomic amplification of these genes been reported.

Breast cancer is the second leading cause of cancer deaths in American women; it is estimated that an American woman has at least a 10% cumulative lifetime risk of developing this disease. Early diagnosis is an important factor in breast cancer prognosis and affects not only survival rate, but also the range of therapeutic options available to the patient. For instance, if diagnosed early, a "lumpectomy" may be performed, whereas later diagnosis tends to be associated with more invasive and traumatic surgical treatments such as radical mastectomy. The treatment of other cancers likewise is benefited by early diagnosis, for instance the prognosis in the treatment of lung cancer, colorectal cancer and prostate cancers is greatly improved by early diagnosis. There is a need for simple and reliable methods of diagnosis of cancers in general and of breast cancer in particular, to improve cancer treatment. Moreover, the identification of proteins involved in cancer (*e.g.*, breast cancer) can lead to better diagnosis and treatment of this increasingly prevalent disease.

SUMMARY

The foregoing problems are addressed by the present disclosure, wherein a strong linkage has been found between genomic amplification of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene and cancer

As disclosed herein, one or more of the specific genes/encoding sequences RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and G29449 is amplified in some human tumors, particularly breast tumors and cell lines derived therefrom. The genomic amplification of these genes, and associated nucleotide sequences, is diagnostic of neoplasia or the potential therefor. Detecting the elevated expression of one or more human RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-encoded products is also diagnostic of neoplasia or the potential for neoplastic transformation. Other genetic alterations leading to elevated/altered RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 expression may be involved in tumorigenesis also, such as mutations in regulatory regions of the gene.

Certain embodiments of the disclosure are methods of detecting a biological condition associated with an abnormal 17q23 locus, for instance by detecting the abnormal 17q23 locus or abnormal protein expression from the locus. Such a locus may be the nucleotide sequence of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene/coding sequence, or complementary sequences, or fragments of these genes and complementary sequences of such fragments. Abnormal protein expression from this 17q23 locus can include expression of proteins encoded for by nucleotides within the locus (*e.g.*, the proteins

encoded for by the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene), as well as expression of fragments of proteins or peptides that normally would not be expressed. Expression of such proteins, fragments and peptides can be from either strand of the 17q23 locus DNA.

Other embodiments of the current disclosure are methods of detecting a biological condition that is associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid or an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein expression, for instance by detecting the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid or the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 expression. Such methods can include *in vitro* amplifying RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid(s), for instance by polymerase chain reaction. Appropriate oligonucleotide primers for such amplification include primers derived from Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and/or G29449, nucleic acid sequences having at least 80% sequence identity with one of these sequences. (Each of these accession numbers are from GenBank except for stSG39547, which designation is a "standard name" applied to this sequence; this is a portion of GenBank Accession No. NT_010740, corresponding to residues 205450-205593 of that accesion.) Usually such oligonucleotides will be 6 or more nucleic acids in length. Further aspects of the disclosure are methods of detecting abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid through nucleotide hybridization (*e.g.*, with a detectable probe) or detection of over-expressed or otherwise aberrant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 mRNA.

Methods according to this disclosure also can detect a biological condition associated with abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 through the detection of abnormal expression of a protein encoded for by RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449, *e.g.* by Western or ELISA. Appropriate protein specific binding agents in such methods include RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific antibodies (*e.g.*, monoclonal and polyclonal) and functional fragments thereof. Such antibodies will recognize an epitope from the amino acid sequence that encodes the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene, amino acid sequences that differ from such sequences by one or more conservative amino acid substitutions; amino acid sequences having at least 60% sequence identity to these sequences; or antigenic fragments of such amino acid sequences.

In other embodiments, the method of detecting the biological condition includes reacting a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 molecule (*e.g.*, a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 encoding nucleic acid or encoded protein) contained in a clinical sample with a reagent such as a corresponding RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-specific binding agent (*e.g.*, a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-specific oligonucleotide or a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein-specific binding agent such as an antibody (monoclonal or polyclonal) or fragment thereof)), to form a RAD51C:agent, S6K:agent, PAT1/APPBP2:agent, TBX2:agent MUL:agent, TRAP240:agent, AA806470:agent, AA635172:agent, stSG39547:agent, and/or G29449:agent complex.

Samples for use in the methods described herein can be taken from one or more cells of the subject. Such cells may include neoplastic cells or tumor cells.

In certain aspects of this disclosure, the methods herein disclosed can be used for presymptomatic screening of an individual for neoplasia or a predisposition to neoplasia.

Biological conditions that can be detected using the provided methods include proliferative diseases and neoplasms, for instance a tumor (especially a breast tumor). In particular embodiments, the tumor is malignant; in others, it is metastatic. Abnormality, in reference to abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid or expression, includes an increase in the amount of a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-specific nucleic acid or protein.

Abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acids may be those that are amplified, for instance by genomic (gene) amplification. Abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 expression may be increased, decreased, deregulated or mis-regulated (as to subcellular or macro-localization, or as to the time of expression).

In various embodiments, appropriate RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid include Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid sequences having at least 80% sequence identity with these sequences; and fragments of these sequences.

Oligonucleotides and nucleic acids as provided herein can be engineered into recombinant DNA vectors. Such vectors can also include a promoter, to which the oligonucleotide or nucleic acid can be operably linked in either the sense or antisense orientation. These vectors can be transformed

into cells or into non-human animals, thereby producing transgenic animals. Such molecules, cells, and non-human transgenic animals are further embodiments of the disclosure.

Methods of modifying the level of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 encoded protein expression in a subject are further encompassed. Such methods include expressing a recombinant genetic construct of the disclosure in a subject. For instance, a construct containing at least 10 consecutive nucleotides of the sequence shown in Accession Number AF029669, NM_003161, NM_006380, or U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and/or G29449, in antisense orientation relative to a promoter, could be expressed in a subject to modify RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 encoded protein expression, respectively, according to the disclosure.

Further aspects are kits for detecting an overabundance of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid or RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 expression. Such kits may comprise oligonucleotides and/or peptides of the disclosure (RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleotide or protein specific binding agents). Some kits provided herein can be used to test a subject for a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid or an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 expression and include a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein binding agent, respectively. Certain of these kits will also include a means for detecting binding of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein binding agent to a corresponding RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 polypeptide. Binding agents in these kits may be any agent that specifically recognizes an epitope on a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein or peptide, such as monoclonal or polyclonal antibodies.

Other kits provided herein detect a genetic mutation in a sample of nucleic acid such as a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-encoding nucleic acid. These kits may include at least one oligonucleotide capable of specifically hybridizing with a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid, PCR amplification primer(s) or hybridization oligonucleotide(s), and a fluorescent (or otherwise) labeled nucleic acid probe that selectively hybridizes to the oligonucleotide. These components can be

disposed in one or more containers. The labeled nucleic acid probe of such kits can have a length of between 5 and 500 nucleotides.

Instructions can optionally be included in any of the kits. Such instructions can provide guidance in comparing the measured amount of test material to a set standard or control sample. They may also provide guidance in interpreting experimental results in order to determine whether the subject tested suffers from a biological condition, for instance one associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying sequence listing.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is series of Northern blots showing the expression levels of S6K, TBX2, PAT1, RAD51C, NACA, and SIGMA1B in various breast cancer cell lines. The cell lines analyzed are indicated above the lanes: 1, MCF-7; 2 ZR-75-1wt; 3, KPL-1; 4, BT-474; 5, SUM-52; 6, MDA-436; 7, HBL-100. The size of each transcript is shown on the left side of the corresponding picture. GAPDH was used as a loading control.

Figure 2 shows expression levels of S6K in various breast cancer cell lines. Figure 2A is a blot from a Northern analysis of S6K mRNA expression in seven breast cancer cell lines. Overexpression of two transcripts (2.5 and 6 kb) is seen in cell lines with high level S6K amplification (BT474, KPL-1, ZR-75-1wt, MCF-7). The membrane was rehybridized with GAPDH as a control. Figure 2B is a Western analysis of proteins in five cell lines, probed with anti-$p70^{S6k}$ antibody. Overexpression of both isoforms (p70 and p85) is seen in cell lines with S6K amplification (MCF-7, KPL-1, ZR-75-1wt). The membrane was re-stained with tubulin antibody as a control. Figure 2C is a graph representing S6K phosphotransferase activity in breast cancer cell lines. Mean + 1 SD of three parallel measurements are shown. Elevated S6K activity is seen in cell lines with amplification.

Figure 3 is a series of graphs that depict cancer-specific survival rates for patients (A) with and without S6K amplification and (B) high or low protein expression according to analysis of a tissue microarray. Figure 3C is a graph showing the combined prognostic significance of S6K and HER-2 amplification. When both genes are amplified, the outcome is worse than when either one of these two genes alone is amplified.

Figure 4A is a physical and transcript map of the 17q23 region. A minimal tiling path consisting of 57 clones between markers WI-16256 and D17S948 is shown. YAC clones are symbolized by thick bars, and BAC, PAC and P1 clones by thin bars. STSs are indicated on the top of the contig while ESTs and known genes are shown under the contig. The map is not drawn in

scale. **Figure 4B** is a graph showing amplicon mapping in MCF7 and BT-474 breast cancer cell lines. DNA sequence copy number changes were analyzed by interphase FISH using 22 clones from the contig (indicated in red in panel A). The mean copy number per cell is plotted for each probe. **Figure 4C** shows a representative interphase FISH with BAC 172h13 and chromosome 17 centromere probes to MCF7 breast cancer cell line, in dark and light grey, respectively.

Figure 5 shows amplicon mapping in primary breast tumors by FISH to tissue microarray. **Figure 5A** is a graph that represents amplification frequencies of 10 clones from the 17q23 contig in 184 primary breast tumors. **Figure 5B** shows amplification patterns in 27 primary breast tumors that displayed amplification with at least one of the probes tested. Rows correspond to individual tumor samples and columns to each clone. White boxes represent no copy number increase and shaded boxes indicate amplification.

Figure 6A and 6B show a copy number and expression survey (respectively) of chromosome 17 specific genes in MCF7 breast cancer cell line by cDNA microarrays. The copy number and expression ratio data are plotted as a function of the position of the clones in the radiation hybrid map. Individual data points are connected with a line. A moving average of three (a mean copy number ratio of three adjacent clones) is shown in **Figure 6A. Figure 6C** shows the copy number ratio as a function of the expression ratio in MCF7 breast cancer cell line. The oval indicates a set of genes that are both highly amplified and overexpressed.

Figure 7 is an expression map of the 17q23 region in six breast cancer cell lines. Each column represents one breast cancer cell line and each row a transcribed sequence arranged according to their physical order in the 17q23 contig from centromere to telomere. ESTs obtained from the genomic sequence are only shown if they were overexpressed in at least three cell lines and are indicated with an asterisk. Shade coding for the expression ratio is shown below the expression map. The locations of the common amplification regions identified in the breast cancer cell lines are indicated with vertical bars to the left of the columns.

DETAILED DESCRIPTION

I. Abbreviations and Terms

A. *Abbreviations*

ASO: allele-specific oligonucleotides

BAC: bacterial artificial chromosome

CGH: comparative genomic hybridization

FISH: fluorescent *in situ* hybridization

ORF: open reading frame

PCR: polymerase chain reaction

RACE: rapid amplification of cDNA ends

RDA: representational difference analysis

SAGE: serial analysis of gene expression

SSCP: single-strand conformational polymorphism

SSRP: simple sequence repeat polymorphism

STS: sequence tagged site

UTR: untranslated regions

YAC: yeast artificial chromosome

B. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments, the following explanation of terms is provided:

Abnormal: Deviation from normal characteristics. Normal characteristics can be found in a control, a standard for a population, etc. For instance, where the abnormal condition is a disease condition, such as neoplasia, a few appropriate sources of normal characteristics might include an individual who is not suffering from the disease (*e.g.*, neoplasia), a population standard of individuals believed not to be suffering from the disease, etc.

Likewise, abnormal may refer to a condition that is associated with a disease. The term "associated with" includes an increased risk of developing the disease as well as the disease itself. For instance, a certain abnormality (such as an abnormality in RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein expression) can be described as being associated with the biological condition of a neoplasm; thus, the abnormality is predictive both of an increased risk of developing a neoplasm and of the presence of a neoplasm. In some instances, more than one of the above-listed proteins may be abnormal in a single cell or sample, and it may be the combination of these abnormalities (such as, in any two of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449, such as both RAD51C and S6K, or in any three of these proteins or any five, or so forth even up to all ten) that is indicative of the biological condition.

An abnormal nucleic acid, such as an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleic acid, is one that is

Antisense, Sense, and Antigene: Double-stranded DNA (dsDNA) has two strands, a 5' -> 3' strand, referred to as the plus strand, and a 3' -> 5' strand (the reverse compliment), referred to as the minus strand. Because RNA polymerase adds nucleic acids in a 5' -> 3' direction, the minus strand of the DNA serves as the template for the RNA during transcription. Thus, the RNA formed will have a sequence complementary to the minus strand and identical to the plus strand (except that U is substituted for T).

Antisense molecules are molecules that are specifically hybridizable or specifically complementary to either RNA or the plus strand of DNA. Sense molecules are molecules that are specifically hybridizable or specifically complementary to the minus strand of DNA. Antigene molecules are either antisense or sense molecules directed to a dsDNA target.

Binding or stable binding: An oligonucleotide binds or stably binds to a target nucleic acid if a sufficient amount of the oligonucleotide forms base pairs or is hybridized to its target nucleic acid, to permit detection of that binding. Binding can be detected by either physical or functional properties of the target:oligonucleotide complex. Binding between a target and an oligonucleotide can be detected by any procedure known to one skilled in the art, including both functional and physical binding assays. Binding may be detected functionally by determining whether binding has an observable effect upon a biosynthetic process such as expression of a gene, DNA replication, transcription, translation and the like.

Physical methods of detecting the binding of complementary strands of DNA or RNA are well known in the art, and include such methods as DNase I or chemical footprinting, gel shift and affinity cleavage assays, Northern blotting, dot blotting and light absorption detection procedures. For example, one method that is widely used, because it is so simple and reliable, involves observing a change in light absorption of a solution containing an oligonucleotide (or an analog) and a target nucleic acid at 220 to 300 nm as the temperature is slowly increased. If the oligonucleotide or analog has bound to its target, there is a sudden increase in absorption at a characteristic temperature as the oligonucleotide (or analog) and target disassociate from each other, or melt.

The binding between an oligomer and its target nucleic acid is frequently characterized by the temperature (T_m) at which 50% of the oligomer is melted from its target. A higher (T_m) means a stronger or more stable complex relative to a complex with a lower (T_m).

Breast cancer: A carcinoma of the breast.

Cancer: A cancer is a biological condition denoting the presence of a malignant tumor, in which a neoplasm has undergone characteristic anaplasia with loss of differentiation, increased rate of growth, invasion of surrounding tissue, and which is capable of metastasis.

The term cancer includes (without limitation) breast carcinomas (*e.g.* lobular and duct carcinomas), and other solid tumors, sarcomas, and carcinomas of the lung like small cell carcinoma, large cell carcinoma, squamous carcinoma, and adenocarcinoma, mesothelioma of the lung, colorectal adenocarcinoma, stomach carcinoma, prostatic adenocarcinoma, ovarian carcinoma such

as serous cystadenocarcinoma and mucinous cystadenocarcinoma, ovarian germ cell tumors, testicular carcinomas, and germ cell tumors, pancreatic adenocarcinoma, biliary adenocarcinoma, heptacellular carcinoma, bladder carcinoma including transitional cell carcinoma, adenocarcinoma, and squamous carcinoma, renal cell adenocarcinoma, endometrial carcinomas including adenocarcinomas and mixed Mullerian tumors (carcinosarcomas), carcinomas of the endocervix, ectocervix, and vagina such as adenocarcinoma and squamous carcinoma, tumors of the skin like squamous cell carcinoma, basal cell carcinoma, melanoma, and skin appendage tumors, esophageal carcinoma, carcinomas of the nasopharynx and oropharynx including squamous carcinoma and adenocarcinomas, salivary gland carcinomas, brain and central nervous system tumors including tumors of glial, neuronal, and meningeal origin, tumors of peripheral nerve, soft tissue sarcomas and sarcomas of bone and cartilage. Particular embodiments refer more specifically to breast tumors and gynecological tumors, such as cervical cancer.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. cDNA may also contain untranslated regions (UTRs) that are responsible for translational control in the corresponding RNA molecule. cDNA is usually synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Comparative genomic hybridization: A technique of differential labeling of test DNA and normal reference DNA, which are hybridized simultaneously to chromosome spreads, as described in Kallioniemi *et al.* (*Science* 258:818-821, 1992), incorporated by reference.

Complementarity and percentage complementarity: Molecules with complementary nucleic acids form a stable duplex or triplex when the strands bind, (hybridize), to each other by forming Watson-Crick, Hoogsteen or reverse Hoogsteen base pairs. Stable binding occurs when an oligonucleotide remains detectably bound to a target nucleic acid sequence under the required conditions.

Complementarity is the degree to which bases in one nucleic acid strand base pair with the bases in a second nucleic acid strand. Complementarity is conveniently described by percentage, *i.e.* the proportion of nucleotides that form base pairs between two strands or within a specific region or domain of two strands. For example, if 10 nucleotides of a 15-nucleotide oligonucleotide form base pairs with a targeted region of a DNA molecule, that oligonucleotide is said to have 66.67% complementarity to the region of DNA targeted.

In the present disclosure, "sufficient complementarity" means that a sufficient number of base pairs exist between the oligonucleotide and the target sequence to achieve detectable binding, and in the case of the binding of an antigen, disrupt expression of gene products (such as one or more of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 genes). When expressed or measured by percentage of base pairs formed, the percentage complementarity that fulfills this goal can range from as little as about 50%

complementarity to full, (100%) complementary. In general, sufficient complementarity is at least about 50%, about 75% complementarity, about 90% or 95% complementarity, and or about 98% or even 100% complementarity.

A thorough treatment of the qualitative and quantitative considerations involved in establishing binding conditions that allow one skilled in the art to design appropriate oligonucleotides for use under the desired conditions is provided by Beltz *et al. Methods Enzymol* 100:266-285, 1983, and by Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

DNA (deoxyribonucleic acid): DNA is a long chain polymer which comprises the genetic material of most living organisms (some viruses have genes comprising ribonucleic acid (RNA)). The repeating units in DNA polymers are four different nucleotides, each of which comprises one of the four bases, adenine, guanine, cytosine and thymine bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid in a polypeptide, or for a stop signal. The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

Unless otherwise specified, any reference to a DNA molecule is intended to include the reverse complement of that DNA molecule. Except where single-strandedness is required by the text herein, DNA molecules, though written to depict only a single strand, encompass both strands of a double-stranded DNA molecule. Thus, a reference to the nucleic acid molecule that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449, or a fragment thereof, encompasses both the sense strand and its reverse complement. Thus, for instance, it is appropriate to generate probes or primers from the reverse complement sequence of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 encoding sequence.

Deletion: The removal of a sequence of DNA, the regions on either side of the removed sequence being joined together.

Gene amplification or genomic amplification: An increase in the copy number of a gene or a fragment or region of a gene or associated 5' or 3' region, as compared to the copy number in normal tissue. An example of a genomic amplification is an increase in the copy number of an oncogene. A "gene deletion" is a deletion of one or more nucleic acids normally present in a gene sequence and, in extreme examples, can include deletions of entire genes or even portions of chromosomes.

Specific levels of genomic amplification can include integer amplifications, for instance a two-fold, three-fold, five-fold or ten fold or greater increase in the number of copies of a particular gene. Within any one sample, however, the number of gene copies in the cells may vary around a mean value. Because of this, non-integer amplifications are also permitted, for instance where

amplification is counted in several (for instance, about 20) cells within a sample and the number of gene copies is averaged across the counted sample. Thus, amplifications of 1.2, 1.5, 1.8, 2.2, etc., may also be indicative of amplification that is associated with a biological condition as discussed herein. In general, higher copy-number amplification, whether integer or non-integer, is indicative of a more severe biological condition, or a less favorable prognosis.

Gene copy number can be determined by any of several well-known techniques, including those discussed herein.

Genomic target sequence: A sequence of nucleotides located in a particular region in the human genome that corresponds to one or more specific genetic abnormalities, such as a nucleotide polymorphism, a deletion, or an amplification. The target can be for instance a coding sequence; it can also be the non-coding strand that corresponds to a coding sequence.

Hybridization: Oligonucleotides and their analogs hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary bases. Generally, nucleic acid consists of nitrogenous bases that are either pyrimidines (cytosine (C), uracil (U), and thymine (T)) or purines (adenine (A) and guanine (G)). These nitrogenous bases form hydrogen bonds between a pyrimidine and a purine, and the bonding of the pyrimidine to the purine is referred to as "base pairing." More specifically, A will hydrogen bond to T or U, and G will bond to C. "Complementary" refers to the base pairing that occurs between to distinct nucleic acid sequences or two distinct regions of the same nucleic acid sequence. For example, a therapeutically effective oligonucleotide can be complementary to a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-encoding mRNA, or an RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-encoding dsDNA.

"Specifically hybridizable" and "specifically complementary" are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between the oligonucleotide (or its analog) and the DNA or RNA target. The oligonucleotide or oligonucleotide analog need not be 100% complementary to its target sequence to be specifically hybridizable. An oligonucleotide or analog is specifically hybridizable when binding of the oligonucleotide or analog to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA, and there is a sufficient degree of complementarity to avoid non-specific binding of the oligonucleotide or analog to non-target sequences under conditions where specific binding is desired, for example under physiological conditions in the case of *in vivo* assays or systems. Such binding is referred to as specific hybridization.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency

of hybridization, though waste times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, chapters 9 and 11, herein incorporated by reference.

For purposes of the present disclosure, "stringent conditions" encompass conditions under which hybridization will only occur if there is less than 25% mismatch between the hybridization molecule and the target sequence. "Stringent conditions" may be broken down into particular levels of stringency for more precise definition. Thus, as used herein, "moderate stringency" conditions are those under which molecules with more than 25% sequence mismatch will not hybridize; conditions of "medium stringency" are those under which molecules with more than 15% mismatch will not hybridize, and conditions of "high stringency" are those under which sequences with more than 10% mismatch will not hybridize. Conditions of "very high stringency" are those under which sequences with more than 6% mismatch will not hybridize.

***In vitro* amplification:** *In vitro* amplification can be performed by any method used to increase the copy number of a nucleic acid molecule. One appropriate technique for amplifying a nucleic acid molecule is the polymerase chain reaction (PCR). PCR is a technique in which cycles of denaturation, annealing with primer oligonucleotide, and extension with DNA polymerase are used to amplify (increase) the number of copies of a target DNA sequence.

Injectable composition: A pharmaceutically acceptable fluid composition including at least one active ingredient, *e.g.*, a therapeutically effective RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific oligonucleotide. The active ingredient is usually dissolved or suspended in a physiologically acceptable carrier, and the composition can additionally include minor amounts of one or more non-toxic auxiliary substances, such as emulsifying agents, preservatives, and pH buffering agents and the like. Such injectable compositions that are useful for use with the nucleotides and proteins of this disclosure are conventional; appropriate formulations are well known in the art.

Isolated: An "isolated" biological component (such as a nucleic acid molecule, protein or organelle) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

Neoplasm: A new and abnormal growth, particularly a new growth of tissue or cells in which the growth is uncontrolled and progressive. A tumor is an example of a neoplasm.

Nucleotide: "Nucleotide" includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base linked to an

amino acid, as in a peptide nucleic acid (PNA). A nucleotide is one monomer in a polynucleotide. A nucleotide sequence refers to the sequence of bases in a polynucleotide.

Oligonucleotide: An oligonucleotide is a plurality of joined nucleotides joined by native phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Particular oligonucleotides and oligonucleotide analogs can include linear sequences up to about 200 nucleotides in length, for example a sequence (such as DNA or RNA) that is at least 6 bases, for example at least 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 or even 200 bases long, or from about 6 to about 50 bases, for example about 10-25 bases, such as 12, 15 or 20 bases.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Open reading frame: A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide.

Ortholog: Two nucleic acid or amino acid sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. Orthologous sequences are also homologous sequences.

Parenteral: Administered outside of the intestine, *e.g.*, not via the alimentary tract. Generally, parenteral formulations are those that will be administered through any possible mode except ingestion. This term especially refers to injections, whether administered intravenously, intrathecally, intramuscularly, intraperitoneally, or subcutaneously, and various surface applications including intranasal, intradermal, and topical application, for instance.

Peptide Nucleic Acid (PNA): An oligonucleotide analog with a backbone comprised of monomers coupled by amide (peptide) bonds, such as amino acid monomers joined by peptide bonds.

Pharmaceutically acceptable carriers: The pharmaceutically acceptable carriers useful in this disclosure are conventional. Martin, *Remington's Pharmaceutical Sciences*, published by Mack Publishing Co., Easton, PA, 19th Edition, 1995, describes compositions and formulations suitable for pharmaceutical delivery of the nucleotides and proteins herein disclosed.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (*e.g.*, powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Probes and primers: Nucleic acid probes and primers can be readily prepared based on the nucleic acid molecules provided in this disclosure, for instance any of the RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, and/or G29449-encoding sequences. It is also appropriate to generate probes and primers based on fragments or portions of these disclosed nucleic acid molecules. Also appropriate are probes and primers specific for the reverse complement of the disclosed sequences, as well as probes and primers to 5' or 3' regions surrounding the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene.

A probe comprises an isolated nucleic acid attached to a detectable label or other reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

Primers are short nucleic acid molecules, for instance DNA oligonucleotides 10 nucleotides or more in length. Longer DNA oligonucleotides may be about 15, 20, 25, 30 or 50 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using nucleic acid probes and primers are described, for example, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989), Ausubel *et al.* (ed.) (In *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998), and Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). One of ordinary skill in the art will appreciate that the

specificity of a particular probe or primer increases with its length. Thus, for example, a primer comprising 30 consecutive nucleotides of a RAD51C-encoding nucleotide (a "RAD51C primer" or "RAD51C probe") will anneal to a target sequence, such as another RAD51C gene homolog from a gene family contained within a human genomic DNA library, with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise at least 20, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 nucleotide sequences.

The disclosure thus includes isolated nucleic acid molecules that comprise specified lengths of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA sequences, especially for use in the detection of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 abnormalities such as amplification. Such molecules may comprise at least 10, 15, 20, 23, 25, 30, 35, 40, 45 or 50 consecutive nucleotides of these sequences or more, and may be obtained from any region of the disclosed sequences. By way of example, the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 sequences may be apportioned into halves or quarters based on sequence length, and the isolated nucleic acid molecules may be derived from the first or second halves of the molecules, or any of the four quarters. The human PAT1/APPBP2 gene can be used to illustrate this. The prototypical human PAT1/APPBP2 shown in GenBank Accession Number NM_006380 is 2385 nucleotides in length and so may be hypothetically divided into about halves (nucleotides 1-1192 and 1193-2385) or about quarters (nucleotides 1-586, 587-1192, 1193-1779 and 1780-2385). The cDNA also could be divided into smaller regions, *e.g.* about eighths, sixteenths, twentieths, fiftieths and so forth, with similar effect.

Another mode of division is to select the 5' (upstream) and/or 3' downstream region associated with any of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNAs or RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genes.

Nucleic acid molecules may be selected that comprise at least 10, 15, 20, 25, 30, 35, 40, 50 or 100 or more consecutive nucleotides of any of these or other portions of the human RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA, gene, and associated flanking regions.

Protein: A biological molecule expressed by a gene and comprised of amino acids.

Purified: The term "purified" does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the protein referred to is more pure than the protein in its natural environment within a cell or within a production reaction chamber (as appropriate).

[redacted]

RAD51C-, S6K, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-associated nucleic acid: These terms encompass the gene or genes (or gene fragments) that encode a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 corresponding protein or a portion thereof, RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 cDNA molecules and the reverse complement thereof (and fragments), RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 mRNA, and 5' and 3' flanking regions of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 cDNA or gene sequence (including regulatory regions and other regions that are abnormal in various biological conditions, including for instance neoplasia), respectively.

RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA: A cDNA molecule which, when transfected or otherwise introduced into cells, expresses the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein, respectively. The RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA can be derived, for instance, by reverse transcription from the mRNA encoded by the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, respectively, and lacks internal non-coding segments and transcription regulatory sequences present in the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene.

RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene: A gene that codes for a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449-encoded protein, respectively, the mutation, genomic amplification or altered expression of which is associated with neoplasms, *e.g.* breast or other cancer. The prototypical RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genes are found on human chromosome 17, and more particularly maps to the 17q23 region of this chromosome and have been disclosed, respectively, and have been assigned Accession Numbers AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449 for instance. (Each of these accession numbers are from GenBank except for stSG39547, which designation is a "standard name" applied to this sequence; this is a portion of GenBank Accession No. NT_010740, corresponding to residues 205450-205593 of that accesion.)

A mutation of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene may include nucleotide sequence changes, additions or deletions, including deletion of large portions or the entire RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, or amplification of all or

substantially all of the gene or a coding sequence thereof. Alternatively, genetic expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 can be deregulated by a mutation such that RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 mRNA and/or protein is over- or under-expressed or expressed in the improper place or at the improper time. Over- or under-expression is compared to a control or a standard for a population. A mutation of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene would be, respectively, an example of an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid.

The term "RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene" is understood to include the various sequence polymorphisms and allelic variations that exist within the population for each of these individual genes (*e.g.*, the term "RAD51C gene" includes polymorphisms and allelic variants of this molecule). This term relates primarily to an isolated coding sequence, but can also include some or all of the flanking regulatory elements and/or intron sequences and encompasses the non-coding as well as the coding strand of the corresponding DNA.

The RNA transcribed from a mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene is mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 messenger RNA (mRNA), respectively.

RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein: The protein encoded for by the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA (as exemplified by the protein sequences shown in Accession Number AF029669, NM_003161, NM_006380, U28049, BAA74921, AAD22031, respectively, or produced through the transcription of the nucleic acid sequences represented by the ESTs AA806470, AA635172, stSG39547, or G29449). This definition is understood to include the various sequence polymorphisms that exist, wherein amino acid substitutions or short insertions or deletions in the protein sequence do not affect the essential functions of the protein.

A RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein encoded for by, respectively, a mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene would be mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein. RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein mutations include alterations in single amino acids within the protein, short or long deletions or amplifications of protein regions. Also encompassed are dis-regulated (loss of normal temporal or spatial regulation) (organ or tissue

specific) and mis-regulated (alteration of normal temporal or spatial regulation) production of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein. A mutation of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein would be an example of a corresponding abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

A "RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific probe" is a probe that binds substantially only to a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein, respectively. Such a protein specific probe includes antibodies and functional fragments thereof, as discussed below.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

Representational difference analysis: A PCR-based subtractive hybridization technique used to identify differences in the mRNA transcripts present in closely related cell lines.

Serial analysis of gene expression: The use of short diagnostic sequence tags to allow the quantitative and simultaneous analysis of a large number of transcripts in tissue, as described in Velculescu *et al.* (*Science* 270:484-487, 1995).

Specific binding agent: An agent that binds substantially only to a defined target. Thus a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein-specific binding agent binds substantially only the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein, respectively. As used herein, the term "RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific binding agent" includes anti-RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein antibodies (and functional fragments thereof) and other agents (such as soluble receptors) that bind substantially only to the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein, respectively.

Polyclonal anti-$p70^{S6k}$ antibody is available commercially, for instance from Santa Cruz Biotechnology (Santa Cruz, CA), product no. sc-230. In addition, anti-RAD51C, anti-S6K, anti-PAT1/APPBP2, anti-TBX2, anti-MUL, anti-TRAP240, anti-AA806470, anti-AA635172, anti-stSG39547, or anti-G29449 protein antibodies may be produced using standard procedures described in a number of texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). The determination that a particular agent binds substantially only to the RAD51C, S6K,

PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein may readily be made by using or adapting routine procedures. One suitable *in vitro* assay makes use of the Western blotting procedure (described in many standard texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988)). Western blotting may be used to determine that a given RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein antibody protein binding agent, such as an anti-protein monoclonal antibody, binds substantially only to the corresponding protein.

Shorter fragments of antibodies can also serve as specific binding agents. For instance, FAbs, Fvs, and single-chain Fvs (SCFvs) that bind to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 would be RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific binding agents, respectively. These antibody fragments are defined as follows: (1) FAb, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) FAb', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two FAb' fragments are obtained per antibody molecule; (3) (FAb')2, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(Ab')2, a dimer of two FAb' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine.

Subject: Living multi-cellular vertebrate organisms, a category that includes both human and non-human mammals

Target sequence: "Target sequence" is a portion of ssDNA, dsDNA or RNA that, upon hybridization to a therapeutically effective oligonucleotide or oligonucleotide analog, results in the inhibition of expression of a specified protein, such as RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein expression. Either an antisense or a sense molecule can be used to target a portion of dsDNA, since both will interfere with the expression of that portion of the dsDNA. The antisense molecule can bind to the plus strand, and the sense molecule can bind to the minus strand. Thus, target sequences can be ssDNA, dsDNA, and RNA.

Therapeutically Effective Oligonucleotides and Oligonucleotide Analogs: An oligonucleotide or analog thereof, capable of inhibiting by at least 15% the expression of a specific target gene. Specific examples of therapeutically effective oligonucleotides and oligonucleotide

analogs are characterized by their ability to inhibit the expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449.

Complete inhibition is not necessary for therapeutic effectiveness. Therapeutically effective oligonucleotides and oligonucleotide analogs are characterized by their ability to inhibit the expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449. Inhibition is defined as any reduction in RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression seen when compared to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 production in the absence of the oligonucleotide or oligonucleotide analog. Additionally, some oligonucleotides will be capable of inhibiting the expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 by at least 15%, 30%, 40%, 50%, 60%, or 70%, or more.

Therapeutically effective oligonucleotides and oligonucleotide analogs are additionally characterized by being sufficiently complementary to RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoding nucleic acid sequences. As described herein, sufficient complementary means that the therapeutically effective oligonucleotide or oligonucleotide analog can specifically disrupt the expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449, and not significantly alter the expression of genes other than RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449.

Transformed: A transformed cell is a cell into which has been introduced a nucleic acid molecule by molecular biology techniques. As used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of naked DNA by electroporation, lipofection, and particle gun acceleration.

Tumor: A neoplasm that may be either malignant or non-malignant. "Tumors of the same tissue type" refers to primary tumors originating in a particular organ (such as breast, prostate, bladder or lung). Tumors of the same tissue type may be divided into tumor of different sub-types (a classic example being bronchogenic carcinomas (lung tumors) which can be an adenocarcinoma, small cell, squamous cell, or large cell tumor). Breast cancers can be divided histologically into scirrhous, infiltrative, papillary, ductal, medullary and lobular.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transformed host cell. A vector may include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector may also include one or more selectable marker genes and other genetic elements known in the art.

Unless otherwise explained herein, all technical and scientific terms used have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

II. Linkage of RAD51C, S6K, PAT1/APPBP2, and TBX2 to Cancer

Materials and Methods

Cell lines:

Breast cancer cell lines BT-474, HBL-100, MCF-7 and MDA-436 were obtained from American Type Culture Collection (ATCC, Manassas, VA) and KPL-1 from the German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). The ZR-75-1wt strain came from Dr. Jeff Moscow (National Cancer Institute, Bethesda, MD) and SUM-52 from Dr. Stephen P. Ethier (University of Michigan, Ann Arbor, MI). Cell lines were grown in recommended culture conditions. Interphase cell preparations from cell lines and normal peripheral blood lymphocytes were done according to routine protocols.

Primary tumors:

Array 1: Three hundred seventy two (372) ethanol-fixed primary breast cancers were obtained from the Institute of Pathology, University of Basel. The tumor samples were reviewed by one pathologist, and included 69.6% ductal, 14% lobular, 2.4% medullary, 1.6% mucinous, 0.8% cribriform, 0.8% tubular, 0.5% papillary carcinomas, 4% ductal carcinoma in situ, and 6.1% of other rare histological subtypes or unclassified. The grade distribution was 24% grade 1, 40% grade 2, and 36% grade 3 The pT stage was pT1 in 29%, pT2 in 54%, pT3 in 9% and pT4 in 8%. The average age of the patients was 60 years (range 28-92 years), with 45% having a node-negative and 55% a node-positive disease.

Array 2: For further analysis of S6K, a second microarray was screened. This array consisted of 612 primary breast cancers from the years 1985-1995, from patients with clinicopathological information including an average of 5.4 years of follow-up. Both tumor cohorts were obtained from the Institute of Pathology, University of Basel. The tumor samples were reviewed by one pathologist and included 73.3% ductal, 13.6% lobular, 3% medullary, 2.6% mucinous, 1.5% cribriform, 1.4% tubular, 1.1% papillary carcinomas, 1.9% ductal carcinoma in situ,

and 1.7% of other rare histological subtypes. The grade distribution was 24% grade 1, 40% grade 2, and 36% grade 3. The pT stage was pT1 in 32%, pT2 in 51%, pT3 in 7%, and pT4 in 10%. The average age of the patients was 60 years (range 26-97 years), with 50% having a node-negative and 50% a node-positive disease.

Tissue microarray construction:

The tissue microarrays were constructed as described previously (Kononen *et al.*, *Nat. Med.* 4:844-847, 1998; see also published International Patent Applications No. PCT/US99/04001 and PCT/US99/04000). Briefly, a representative tumor area was marked from hematoxylin-eosin stained sections of each tumor. The blocks and the corresponding histological slides were overlayed for tissue microarray sampling. A tissue microarray instrument (Beecher Instruments, Silver Spring, MD) was used to create holes in a recipient paraffin block, to obtain cylindrical tissue biopsies with a diameter of 0.6 mm from the donor paraffin blocks, and to transfer these biopsies to the recipient block at defined array positions. Multiple 5 µm sections were cut from the tissue microarray block using a microtome with an adhesive-coated tape sectioning system (Instrumedics, Hackensack, NJ).

DNA probes for FISH:

Gene specific BAC clones were obtained by screening a human BAC library (Genome Systems, St. Louis, MO, USA) using polymerase chain reaction with primers specific for the genes. BAC probes were labeled with SpectrumOrange (Vysis Inc., Downers Grove, IL) using random priming. The PAC probe (186o9) specific for S6K was also labeled. SpectrumGreen labeled chromosome 17 centromere probe (Vysis Inc., Downers Grove, IL) was used as a reference.

Copy number analysis by FISH:

Interphase FISH to breast cancer cell lines was done as previously described (Bärlund *et al.*, *Genes Chromo. Cancer*, 20: 372-376, 1997). The hybridizations were evaluated using a Zeiss fluorescence microscope and approximately 20 non-overlapping nuclei with intact morphology based on the DAPI counterstain were scored to determine the mean number of hybridization signals for each test and reference probe

For the tissue microarrays, the FISH protocol was changed according to the type of fixation used. The ethanol-fixed tissue microarray (Array 1) was deparaffinized, denatured at 73° C for 5 minutes in 70% formamide/2 x SSC and dehydrated in an ethanol series (70, 80, and 100%) as previously described (Kononen *et al.*, *Nat. Med.*, 4: 844-847, 1998). The formalin-fixed tissue microarray (Array 2) was deparaffinized, immersed in 0.2 N HCl, incubated in 1M sodium thiocyanate solution at 80°C for 30 minutes, and immersed in a protease solution (0.5 mg/ml in 0.9% NaCl; Vysis Inc., Downers Grove, IL) for 10 minutes at 37° C. These slides were then postfixed in 10 % buffered formalin, air dried, denatured at 73° C for 5 minutes in 70 % formamide/2 x SSC,

dehydrated in an ethanol series, incubated in 4 μg/ml proteinase K at 37° C for 7 minutes, and dehydrated again as above.

Both ethanol-fixed and formalin-fixed slides were hybridized with specific test probes and the chromosome 17 centromere probe overnight at 37° C. An adjacent section from the tissue microarray was hybridized with a SpectrumOrange labeled HER-2 probe (Vysis Inc., Downers Grove, IL). The hybridizations were evaluated using a Zeiss fluorescence microscope. The specimens containing tight clusters of signals or > 3-fold increase in the number of gene-specific probe signals as compared to chromosome 17 centromere in at least 10% of the tumor cells were considered as amplified.

Northern hybridization:

Total RNA was extracted from breast cancer cell lines and Northern hybridization performed using standard methods. Briefly, 10 μg of total RNA was transferred on a Nytran membrane. The blot was prehybridized for 1 hour at 68° C in Express Hybridization solution (CLONTECH, Palo Alto, CA) together with boiled sheared DNA (10 μg/ml, Research Genetics, Huntsville, AL). Sequence verified cDNA inserts (S6K insert, GenBank: H04661; TBX2 insert, GenBank: AI244102) or approximately 900 bp PCR products (RAD51C and PAT1/APPBP2) were ^{32}P-labeled by random priming (Prime-It, Stratagene, La Jolla, CA). The blot was prehybridized for 1 hour at 68° C in Express Hybridization solution (CLONTECH) together with boiled sheared DNA (10μg/ml, Research Genetics, Huntsville, AL). Hybridization was done in the prehybridization solution at 68° C overnight. Blots were washed several times with 2 x SSC/0.05% SDS at 30° C and then in 0.1 x SSC/0.1% SDS at 50° C. In the case of S6K, blots were washed with 2 x SSC/1% SDS at 65° C and then in 0.1 x SSC/0.5% SDS at 55° C. Hybridized probe was detected autoradiographically or by using a Molecular Dynamics PhosphorImager. After stripping, membranes were rehybridized with GAPDH probe to confirm equal loading among samples.

Fluorescent cDNA microarray analysis of MCF-7:

A cDNA microarray was used to screen for up-regulated genes/ESTs in the breast cancer cell line MCF-7, previously shown to have high level amplification at 17q23.

A total of 4209 cDNA clones were printed on glass slides as described (DeRisi *et al.*, *Nat. Genet.*, 14: 457-60, 1996). These clones were part of the 15k set of image consortium cDNA clones described earlier (Duggan *et al.*, *Nat. Genet.*, 21: 10-4, 1999; Ermolaeva *et al.*, *Nat. Genet.*, 20: 19-23, 1998) and 50% of them had radiation hybrid mapping information. The set included 135 clones from chromosome 17 and 8 clones from 17q23. Total RNA was extracted from MCF-7 cells using the Rneasy™ kit (Qiagen, Valencia, CA). The labeling and hybridization was done as previously described (Khan *et al.*, *Cancer Res.* 58: 5009-13, 1998). Briefly, fluorescently labeled cDNAs were prepared from 100-200 μg of total MCF-7 and normal mammary gland RNA (CLONTECH, Palo

Alto, CA) by oligo dT-primed polymerization using SuperScript II™ (SS II) reverse transcriptase (LTI Inc.). The Cy5 labeled MCF-7 cDNA and Cy3 labeled normal mammary cDNA were combined with 8 µg poly(dA) (Pharmacia, Bridgewater, NJ), 4 µg *Escherichia coli* tRNA (Sigma Chemical Co., St. Louis, MO), and 10 µg Cot-1 DNA (Life Technologies, Inc., Rockville, MD) in 0.15% SDS, 3 x SSC. The probe mix was incubated at 98° C for 2 minutes and at 4° C for 10 seconds, and applied onto the microarray slide. The hybridization was carried out at 65° C for 16 hours. The slides were washed in 0.5 x SSC and 0.01% SDS at 55° C for 2 minutes each.

Fluorescence intensities at the targets were measured using a custom-designed laser confocal microscope equipped with a scanning stage, appropriate excitation and emission filters, and a photomultiplier tube detector (Duggan *et al.*, *Nat. Genet.*, 21: 10-4, 1999). Intensity data were integrated over 225-μm^2 pixels and recorded at 16 bits. The two fluorescent images were scanned separately and the color images were formed by arbitrarily assigning tumor intensity values into the red channel and control intensity into the green channel.

Analysis of cDNA microarray data:

The results of the cDNA microarray experiments were analyzed using the ArraySuite program, developed at NHGRI by Yidong Chen, based on the IPLab Spectrum platform (Chen *et al.*, *Biomed, Optics*, 2: 364-74, 1997). The program identifies and segments spots in the two images corresponding to the test (red) and control (green) cDNA hybridizations. After background subtraction, average intensities at each spot in the test hybridization were divided by the average intensity of the same spot in the normal reference hybridization. The ratios were normalized based on the distribution of ratios for 88 house-keeping genes (Khan *et al.*, *Cancer Res.* 58: 5009-13, 1998). This ensured that the ratios obtained were not affected by differential labeling or hybridization efficiency of the two cDNAs, nor by the quantity of the labeled cDNA in the reaction or the photomultiplier voltage settings used in the scanning of the slides by the confocal laser scanner. The ArraySuite program enables analysis of the quantitative ratios at each spot, as well as the visualization of the image for each spot. For the present study, the spot images were displayed by the chromosomal location of these genes. This was done to obtain a visual comparison between DNA amplification sites and chromosomal locations of over-expressed genes. The chromosomal locations of these genes came from radiation hybrid mapping data available in the Human Gene Map 98 (http://www.ncbi.nlm.nih.gov/genemap/).

Western Hybridization:

Breast cancer cell line pellets were lysed in RIPA buffer by repeated passage through a 21 gauge needle and microcentrifuged at 10,000 x g for 10 minutes at 4° C. Approximately 40 µg of total protein was boiled together with loading buffer for 5 minutes and electrophoresed through 4-12% polyacrylamide gradient Tris-Glycine gel (Novex, San Diego, CA), and electroblotted to a

nitrocellulose membrane (Novex, San Diego, CA). Detection of protein was performed by BM Chemiluminescence Western Blotting Kit (Mouse/Rabbit) (Boehringer Mannheim, Indianapolis, IN). Briefly, the membrane was incubated for 1 hour in 1% blocking solution and for 1 hour in primary antibody (anti-$p70^{S6k}$ antibody, Santa Cruz Biotechnology, Santa Cruz, CA) in 0.5% blocking solution at 4° C. The membrane was washed two times in TBS-Tween®20 for 10 minutes each, then twice with 0.5% blocking solution, incubated for 30 minutes at 4° C with peroxidase-labeled secondary antibody (mouse anti-rabbit IgG) and then washed four times with large volumes of TBS-Tween®20 for 15 minutes each. Signal detection by horseradish peroxidase chemiluminescent reaction was done according to the manufacturer's instructions. The membrane was re-stained with tubulin antibody to confirm equal loading among samples.

S6 kinase activity:

The breast cancer cell line monolayers in 100-mm dishes were washed once in PBS, and the cells were lysed by incubating the cultures for 20 minutes in 1 ml of cold modified radioimmunoprecipitation assay buffer (50 mM Tris-HCl, pH 7.4, 1% Nonidet P-40, 0.25 % sodium deoxycholate, 150 mM NaCl, 1 mM EGTA, 1 mM PMSF, 1 mM Na_3VO_4, 1 mM NaF, complete protease inhibitor cocktail (Boehringer Mannheim, Indianapolis, IN)). The lysates were passaged through a 21 gauge needle and centrifuged at 10,000 g for 10 minutes at 4° C. Approximately 200 µg of protein was incubated with 1 µg of anti-$p70^{S6k}$ antibody (Santa Cruz Biotechnology, Santa Cruz, CA) overnight at 4° C. The immunocomplex was captured with 50 µl protein A-agarose for 3 hours at 4° C. The beads were washed four times with cold PBS. The S6K enzyme activity was assayed using the S6 kinase assay kit (Upstate Biotechnology, Lake Placid, NY) according to the manufacturer's recommendations. Briefly, the beads were resuspended in 20 µl of assay dilution buffer, 10 µl of substrate, 10 µl of inhibitor mixture and 10 µl of [γ^{32}P]ATP mixture (75 mM $MgCl_2$, 500 µM ATP, 10 µCi of [γ^{32}P]ATP) and incubated for 10 minutes at 30° C. Aliquots (25 µl) were spotted onto P81 phosphocellulose paper squares and washed three times with 0.75% phosphoric acid and once in acetone. The squares were transferred to scintillation vials and counted in a LS6500 scintillation counter (Beckmann Coulter Inc., Fullerton, CA).

Growth inhibition by rapamycin:

For growth inhibition experiments, 10^4 cells were plated into 96-well microtiter plates and permitted to grow over night. The next day, serial dilutions of rapamycin (Calbiochem, La Jolla, CA) were added and incubated for 70 hours. After the drug removal, the cells were allowed to recover overnight in appropriate culture media and the growth inhibition was measured using the colorimetric MTT test (Chemicon International Inc., Temecula, CA). All experiments were

performed in triplicates and the results were expressed as a percentage inhibition relative to control cultures without rapamycin.

Immunohistochemistry:

Standard indirect immunoperoxidase protocol with diaminobenzidine as a chromogen was used for immunohistochemistry (ABC-Elite, Vector Laboratories, Burlingame, CA). Polyclonal anti-$p70^{S6k}$ antibody (Santa Cruz Biotechnology) was used for detection of S6K (1:5000 dilution). High-temperature (20 minutes in a pressure cooker) antigen retrieval (antigen unmasking solution, Vector Laboratories) procedure was used to enhance the staining. The primary antibody was omitted for negative controls. The S6K staining was subjectively graded into 4 groups: negative (no staining), weak, moderate, or strong cytoplasmic staining. For statistical analyses, the data was combined into two groups. Negative or weak staining was considered low expression and moderate or strong staining high expression.

Statistical analyses:

The correlation of S6K amplification to clinicopathological parameters and the association between S6K amplification and immunohistochemical staining were analyzed using the two-sided Fisher's exact test. The prognostic significance of S6K and HER-2 amplification was analyzed using the Logrank test.

Results

Identification of Specific Gene Amplification in Breast Cancer Cell Lines

Copy number changes of six genes (RAD51C, S6K, SIGMA1B, PAT1, NACA, and TBX2) located at 17q23 were studied by FISH in seven breast cancer cell lines. Four cell lines (BT-474, KPL-1, MCF-7 and ZR-75-1wt) were previously known to have amplification or gain at 17q22-q24 by CGH whereas three cell lines (SUM-52, HBL-100, and MDA-436) did not show any copy number increase at this region (Bärlund *et al.*, *Genes Chromo. Cancer*, 20: 372-376, 1997). All six genes were found to be highly amplified (8- to 19-fold relative to chromosome 17 centromere) in three cell lines (KPL-1, MCF-7, and ZR-75-1wt) (Table 1). In addition, 4- to 5-fold amplification of S6K, SIGMA1B, and PAT1 was seen in BT-474 as well as NACA and TBX2 in SUM-52 (Table 1).

Table 1. Amplification levels of six 17q23 genes relative to chromosome 17 centromere in breast cancer cell lines by FISH.

Cell line	RAD51C	S6K	SIGMA1B	PAT1	NACA	TBX2
MCF-7	18.3	11.7	13	10	16.7	10
ZR-75-1wt	19.2	10.8	11.2	7.7	19.2	11.5
KPL-1	15	10.5	15.8	13.3	16.7	13.3
BT-474	1.8	5.3	3.9	5.4	1.4	1.2
SUM-52	ND	2.4	1.9	1.4	4.9	4.3
HBL-100	1	1	1	1.3	1	1.3
MDA-436	1	1	1	1.5	1	1.5

ND, not determined

To determine whether the amplification led to elevated expression of these genes, Northern analysis was performed (Figure 1). NACA was ubiquitously expressed in all cell lines and SIGMA1B showed elevated expression only in BT-474 and HBL-100 indicating that the expression of these two genes did not correlate with amplification levels. The expression of RAD51C, S6K, and PAT1 and was elevated in all cell lines with amplification (Figure 1). Interestingly, RAD51C was also over-expressed in cell line BT-474, which did not show amplification of this gene. TBX2 was expressed in the three cell lines with high level amplification but it's expression was not detectable in the SUM-52 cell line with low level amplification.

To survey whether amplifications of these four genes (RAD51C, S6K, PAT1, and TBX2) also occur in primary breast cancers, we used FISH analysis on a tumor tissue microarray containing 372 primary breast tumors. Thirty-nine tumors (10.5 %) showed amplification of at least one of the genes. RAD51C was amplified least frequently, in 3.1 % of cases, and it was always co-amplified with the other three genes. S6K amplification was seen in 10.2 % of the tumors, PAT1 in 8.9 %, and TBX2 in 8.6 %. In majority of the cases (62 %), S6K, PAT1, and TBX2 were simultaneously amplified. However, amplification of a single gene was observed in 12 cases (31 %), S6K in five tumors, PAT1 in four and TBX2 in three cases.

cDNA Microarray Analysis Demonstrates that S6K is Most Highly Overexpressed 17q23 Linked Gene

One of the most prominent amplification sites in the MCF-7 cells by CGH is localized to 17q23. To better visualize the correlation between increased copy number and increased expression levels of genes along chromosome 17, the relative expression ratios of all of the genes in the cDNA microarray were arranged according to their chromosomal position. The distal 17q peak in the cDNA microarray profile corresponds to an EST (Image clone ID 133432). This EST showed the highest red to green ratio (8.4) of all evaluable spots on this MCF-7 cDNA microarray experiment. This EST belongs to a Unigene cluster Hs.124943 (http://www.ncbi.nlm.nih.gov/UniGene), located

at 17q23 (radiation hybrid mapping value 369 cR_{3000}). Subsequent sequencing and extension of the cDNA clones in this Unigene cluster indicated that this Unigene represented the 3' UTR of the ribosomal protein S6 kinase gene (S6K). The role of S6K as a possible amplification target gene was therefore explored further.

S6K copy number, expression and enzymatic activity in breast cancer cell lines

S6K gene copy number was analyzed in seven breast cancer cell lines by interphase FISH using a PAC probe (186o9). Five cell lines showed an increase in S6K copy number relative to the chromosome 17 centromere. Four of these had high-level amplification of S6K (BT-474 >5-fold, MCF-7, KPL-1 and ZR-75-1 >10-fold). SUM52 showed only a 2-fold amplification, whereas HBL-100 and MDA-436 had no increase in S6K copy number.

The S6K mRNA and protein expression was studied by Northern and Western analyses. Northern revealed two transcripts (2.5 and 6 kb, Figure 2A), consistent with the literature (Reinhard *et al., Proc Natl. Acad. Sci USA* 89: 4052-56, 1992). Both of the transcripts were highly expressed (8- to 20-fold) in cell lines with high level amplification relative to non-amplified cell lines. Western blotting revealed two bands (70 kD and 85 kD) corresponding to the two known isoforms of S6K protein (Grove *et al., Mol. Cell Biol.*, 11: 5541-50, 1991). Increased expression of both isoforms was seen in the cell lines with S6K amplification (Figure 2B).

The S6K protein is activated by phosphorylation on multiple sites (Chou *et al., Curr. Opinion Cell Biol.*, 7: 806-14, 1995; Pullen *et al., Science*, 279: 707-10, 1998). We therefore studied if the increased expression of S6K protein would also lead to increased S6K activity. The enzymatic activity of S6K was measured in lysates from six breast cancer cell lines. The cell lines with S6K amplification and overexpression showed 3- to 14-fold higher S6K activity than those cell lines with no amplification and overexpression (Figure 2C). The growth inhibitory effects of rapamycin, which blocks the phosphorylation of S6K (Chou *et al., Curr. Opinion Cell Biol.*, 7: 806-14, 1995), was studied using the MTT test. All cell lines showed a concentration-dependent growth inhibition with rapamycin regardless of the amplification status.

S6K Analysis: Tissue microarrays and Clinicopathological Correlations

We applied FISH analysis on tumor tissue microarrays to evaluate the amplification of the S6K gene in vivo, in 668 primary breast tumors. Fifty-nine cases (8.8%) showed amplification of S6K. The S6K protein expression was analyzed in a subset of 445 tumors using immunohistochemistry. Moderate or strong cytoplasmic staining was seen in 69 primary tumors (15.5%). There was a significant association between S6K amplification and high S6K expression ($p=0.0004$, Fisher's exact test) with 41% of the amplified tumors exhibiting high S6K expression.

No significant association was seen between S6K amplification and the age of the patient, tumor size, nodal status, histological grade or the presence of estrogen or progesterone receptors.

However, patients with S6K amplification had a significantly worse survival than patients without S6K amplification (p=0.0021, Figure 3A). Similar prognostic correlation was also found with high S6K expression (p=0.0082, Figure 4B). Since HER-2 oncogene (located on the same chromosome at 17q12) amplification is also associated with poor prognosis in breast cancer (Ross and Fletcher, *Sem. Cancer Biol.*, 9: 125-38, 1999), we analyzed whether S6K showed prognostic significance irrespective of HER-2. An adjacent section from the tissue microarray was hybridized with a probe for HER-2, which showed amplification in 106 cases (17.4%) and a clear prognostic association (p=0.0001). Amplification of both S6K and HER-2 was seen in 27 cases (4.4%) indicating that in more than half of the cases S6K amplification occurred without HER-2 amplification. Furthermore, patients with tumors showing both S6K and HER-2 amplification had a significantly worse prognosis than those with amplification of only one of these genes (p=0.0001, Figure 3C).

III. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer, and linkage of MUL, TRAP240, AA806470, AA635172, stSG39547, and G29449 to cancer

To date, five genes *(RPS6KB1, RAD51C, PATI/APPBP2, SIGMA1B,* and *TBX2)* have been implicated as putative target genes for the 17q23 amplification (Section II, above, and Bärlund *et al.*, *J. Natl. Cancer Inst.* 92:1252-1259, 2000; Couch *et al.*, *Cancer Res.* 59:1408-1411, 1999; Bärlund *et al.*, *Cancer Res.* 60:5340-5344, 2000; Wu *et al.*, *Cancer Res.* 60:5371-5375, 2000). This Section describes the comprehensive characterization of 17q23 amplification in breast cancer, including molecular cloning of an about 4 Mb region at 17q23 and detailed evaluation of the amplicon structure in breast cancer cell lines and primary breast tumors. Full expression profiling of the amplified region was performed using a custom-made cDNA microarray containing 156 transcripts from the 17q23 region as well as all known genes mapping to chromosome 17.

The biological significance of DNA amplification in cancer is thought to be due to the selection of increased expression of a single or few important genes. However, systematic surveys of the copy number and expression of all genes within an amplified region of the genome have not been performed. Here a combination of molecular, genomic, and microarray technologies was used to identify target genes for 17q23, a common region of amplification in breast cancers with poor prognosis. Construction of a 4 Mb genomic contig made it possible to define two common regions of amplification in breast cancer cell lines and 372 primary breast tumors using fluorescence in situ hybridization on tissue microarrays.

Based on the GeneMap'99 information, 17 known genes and 26 ESTs were localized to the contig and analysis of genomic sequence identified 77 additional transcripts. A comprehensive analysis of the expression level of these transcripts in six breast cancer cell lines was carried out using cDNA microarrays. A limited number of consistently overexpressed genes were identified. Of

these, *RPS6KB1*, *MUL*, *APPBP2*, and *TRAP240* as well as one uncharacterized expressed sequence tag were located in the common amplified region. In summary, comprehensive analysis of the 17q23 amplicon revealed a limited number of highly expressed genes that appear to contribute to the more aggressive clinical course observed in breast cancer patients with 17q23 amplified tumors. In particular, six additional amplification targets are now identified, beyond those discussed above in Section II. Those six targets are as follows: two additional previously known genes (MUL, GenBank Accession No. AB020705; and TRAP240, GenBank Accession No. AF117754), and four ESTs with no identified function (Accession Nos. AA806470, AA635172, stSG39547, and G29449).

Materials and Methods

Physical and transcript mapping:

YAC clones representing the 17q23 amplified region were identified from the Whitehead Institute database (available on the internet at http://www-genome.wi.mit.edu/) using mapping information from our previous study (Bärlund *et al.*, *Cytogenet. Cell Genet.* 82:189-191, 1998). Corresponding STSs and ESTs were identified from the dbSTS and dbEST databases (available on the internet at the NCBI Website, http://www.ncbi.nlm.nih.gov/) and their location in the YACs was confirmed by PCR. The STSs and ESTs were used to screen commercially available P1 and BAC libraries (Genome Systems Inc., St. Louis, MO; Research Genetics Inc., Huntsville, AL) or a PAC library provided by Dr. Pieter J. De Jong (Roswell Park Institute, Buffalo, NY). New STSs were generated by clone end sequencing. The STS, EST, and clone end sequences were compared against the nr and htgs databases using the blastn program (available at the NCBI Website, http://www.ncbi.nlm.nih.gov/BLAST/) to identify genomic clones with sequence information. All clones were hybridized to normal metaphase chromosomes to verify their chromosomal location and in some instances fiber-FISH was performed to confirm the clone order as previously described (Bärlund *et al.*, *Cytogenet. Cell Genet.* 82:189-191, 1998; Heiskanen *et al.*, *Biotechniques* 17:928-929, 1994). To create a transcript map for the amplified region, ESTs identified from the Whitehead YAC map and the GeneMap'99 (http://www.ncbi.nlm.nih.gov/genemap) were localized to the contig by PCR.

Breast cancer cell lines:

Seven breast cancer cell lines (BT-474, HCC1428, MCF7, MDA-157, MDA-361, MDA-436, and UACC-893; American Type Culture Collection, Manassas, VA) were used in this study. Normal human mammary epithelial cells (HMEC) were obtained from Clonetics (Walkersville, MD). Cells were grown under recommended culture conditions.

Tissue microarray:

The tissue microarray used in this study has been described previously (Kononen *et al.*, *Nat.*

Med. 4:844-847, 1998) and included 372 ethanol-fixed primary breast cancers from the Institute of Pathology, University of Basel. The tumors were 69.6% ductal, 14% lobular, 2.4% medullary, 1.6% mucinous, 8.4% other rare histological subtypes, and 4% ductal carcinoma in situ. The grade distribution was 24% grade 1, 40% grade 2, and 36% grade 3. All specimens evaluated were anonymous, archival tissue specimens.

Fluorescence in situ hybridization:

Dual-color interphase FISH to breast cancer cell lines was done as previously described (Bärlund *et al.*, *Cytogenet. Cell Genet.* 82:189-191, 1998). Probes were labeled with SpectrumOrange (Vysis, Inc., Downers Grove, IL) by random priming and a SpectrumGreen labeled chromosome 17 centromere probe (Vysis, Inc.) was used as a reference. Twenty-five non-overlapping nuclei with intact morphology based on the 4',6-diamidino-2- phenylindole counterstain were scored to determine the mean number of hybridization signals for each probe. FISH to the tissue microarray was done as described previously (Kononen *et al.*, *Nat. Med.* 4:844-847, 1998). Briefly, consecutive sections of the array were deparaffinized, dehydrated in ethanol, denatured at 74°C for five minutes in 70% formamide/2 x SSC, and hybridized with labeled test and control probes. Specimens containing tight clusters of test probe signals or more than a 3-fold increase in the number of test probe signals, as compared to chromosome 17 centromere signals, in at least 10% of the tumor cells were considered to be amplified.

Chromosome 17 specific cDNA microarray:

Clones for the cDNA microarray were selected based on the information available in the GeneMap'99 (on the internet at the NCBI Website, http://www.ncbi.nlm.nih.gov/genemap) and included all known genes on chromosome 17 and all ESTs from intervals D17S933-D17S930 (293-325 cR, the ERBB2 locus) and D17S791-D17S795 (333-435 cR, the 17q23-q24 region). Seventy-seven additional transcribed sequences (34 Unigene clusters and 43 single ESTs) were identified from genomic sequence of nine BAC clones (1073f15, 178c3, 217l10, 215p18, 987k16, 15k2, 332h18, 136h19, and 264b14) by performing sequence similarity searches to the dbEST database using the GeneMachine software (available at the NHGRI Website, http://genemachine.nhgri.nih.gov/). A total of 636 clones, including 156 clones from the 17q23 region (35 clones from the contig, 77 clones identified from the genomic sequence, and 44 additional transcripts from radiation hybrid map interval 367-386 cR corresponding to our contig), were placed on the array. The preparation and printing of the cDNA clones on glass slides were performed as described previously (DeRisi *et al.*, *Nat. Genet.* 14:457-460, 1996; Shalon *et al.*, *Genome Res.* 6:639-645, 1996).

Copy number and expression analyses by cDNA microarrays:

The CGH microarray analysis was done as previously described (Pollack *et al. Nat. Genet.* 23:41-46, 1999) with slight modifications. Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal placental reference was digested for 14-18 hours with Alu I and Rsa I restriction enzymes (Life Technologies, Inc.) and purified by phenol/chloroform extraction. Six micrograms of digested cell line DNA was labeled with Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, NJ) and 8 μg of placental DNA with Cy5-dUTP (Amersham Pharmacia Biotech) using Bioprime Labeling kit (Life Technologies, Inc.). Hybridization was done according to the protocol by Pollack et al. (*Nat. Genet.* 23:41-46, 1999) and post-hybridization washes as described (Khan *et al.*, *Cancer Res.* 58:5009- 5013, 1998; Mousses *et al.*, in *Functional Genomics*, eds. Livesey. & Hunt (Oxford University Press, Oxford), pp. 113-137, 2000). For the expression analyses, total RNA (MDA-361 and UACC-893) or mRNA (BT-474, HCC1428, MCF7, MDA-157, MDA-436, and HMEC) was extracted by using the FastTrack 2.0 kit (Invitrogen, Carlsbad, CA). The breast cancer cell line MDA-436 with no copy number increase at 17q23 (see results) was used as a standard reference in all experiments. Sixteen micrograms of MDA-436 mRNA was labeled with Cy5-dUTP and 4 μg of test mRNA or 80 μg of total RNA with Cy3-dUTP by use of oligo(dT)-primed polymerization by SuperScript II reverse transcriptase (Life Technologies, Inc., Rockville, MD). The labeled cDNAs were hybridized on microarrays as described previously (Khan *et al.*, *Cancer Res.* 58:5009- 5013, 1998; Mousses *et al.*, in *Functional Genomics*, eds. Livesey, F. J. & Hunt, S. P. (Oxford University Press, Oxford), pp. 113-137, 2000). For both the copy number and expression analyses, the fluorescence intensities at the targets were measured by using a laser confocal scanner (Hewlett Packard, Palo Alto, CA). The fluorescent images from the test and control hybridizations were scanned separately and the data was analyzed using the DEARRAY software (Chen *et al.*, *J. Biomed .Optics* 2:364-374, 1997).

Results

Molecular cloning of the amplified region at 17q23:

Seven YAC clones were determined to be non-chimeric by metaphase FISH and amplified in MCF7 breast cancer cell line. These clones served as the basis for contig construction. Approximately 4 Mb contig (with a single gap of unknown size) was constructed between markers WI-16256 and D17S948 (Fig. 4A). The contig consists of 74 large insert size PAC, BAC, and P1 clones representing at least a two-fold coverage and part of it has been published previously (Paavola *et al.*, *Genome Res.* 9:267-276, 1999). A total of 17 known genes and 26 ESTs were positioned in the contig by STS mapping.

Copy number analysis in breast cancer cell lines and primary breast tumors:

Seven breast cancer cell lines (BT-474, HCC1428, MCF7, MDA-157, MDA-361, MDA-

the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein band (which will be localized at a given position on the gel determined by its molecular weight). Non-specific binding of the antibody to other proteins may occur and may be detectable as a weak signal on the Western blot. The non-specific nature of this binding will be recognized by one skilled in the art by the weak signal obtained on the Western blot relative to the strong primary signal arising from the specific antibody-target protein binding, be the target RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449.

Substantially pure RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein or protein fragment (peptide) suitable for use as an immunogen may be isolated using any art-known techniques, including those discussed above. Concentration of protein or peptide in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per milliliter. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein (or peptides derived from the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus) identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (*Nature* 256:495-497, 1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess un-fused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (*Meth. Enzymol.* 70:419-439, 1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988).

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein (Example 4), which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is

436, and UACC-893) were screened by interphase FISH to define a minimal common region of amplification at 17q23. Copy number analysis using 22 clones from the contig showed high-level amplification in BT-474 and MCF7 cell lines (Fig. 4B and C). Both cell lines had two separate but partly overlapping regions of amplification that allowed identification of two major common regions of involvement: one from clone 242o20 to 58p18 and another from clone 52i20 to 215p18. Clone 52a23 at the distal end of the contig also showed amplification in both cell lines. Low-level amplification of 17q23 was detected in the MDA-157, MDA-361, HCC1428, and UACC-893 cell lines. MDA-157 had an average of 7 copies per cell (2-fold amplification relative to the chromosome 17 centromere) of all 22 probes tested, MDA-361 had 7-12 copies per cell (2- to 3-fold amplification), and HCC1428 and UACC-893 had 6-8 copies per cell (3- to 4-fold amplification). No copy number increase was observed in MDA-436 cell line.

To extend the initial studies on breast cancer cell lines to a very large number of primary tumors, copy numbers of 10 selected clones were analyzed using a tissue microarray containing 372 primary breast tumors. Evaluable results with all 10 clones were obtained in 184 tumors and the amplification frequencies ranged from 3.8% to 12.5% (Fig. 5A) with the most frequently amplified core region (5 clones) located in the middle of our 17q23 map. Twenty-seven tumors (15 %) showed amplification of at least one of the clones (Fig. 5B). The majority of these (20/27; 74%) showed co-amplification of at least five of the clones indicating that amplicons in this region are usually large and continuous (Fig. 5B). The remaining seven tumors showed amplification of one to four clones. The highest frequency of amplification was around clone 224b10, but the differences between the individual clones were not sufficient to confine amplification within the core region defined by the five clones (Fig. 5B).

Comprehensive analysis of copy number and expression of genes from chromosomal region 17q23 using cDNA microarrays:

After the region of amplification was approximately defined by physical mapping and tissue microarray based amplicon mapping, we resorted to high-resolution DNA microarray analysis (a total of 636 cDNA clones from chromosome 17, 156 of these from 17q23) to identify the specific genes involved in this region. The copy number analysis in six breast cancer cell lines (BT-474, HCC1428, MCF7, MDA-157, MDA-361, and UACC-893) by CGH to the cDNA microarray identified a distinct region of increased copy number at 370-390 cR (Fig. 6A). The location of this region corresponds to the location obtained by FISH mapping at 17q23. This broad survey of the entire chromosome indicates that the region defined by us is the major amplification locus at 17q23. The cDNA microarray analysis for expression of the 636 chromosome 17 specific genes indicated multiple highly overexpressed genes corresponding to the amplicon (Fig. 6B and C). Focused evaluation of 156 clones from the 17q23 region in the six cell lines with 17q23 amplification identified a total of 19 transcripts that were overexpressed (greater than 3-fold) in at least three of the

cell lines but not in the normal mammary epithelial cells. Sixteen of the consistently overexpressed sequences were located within our contig (five of these originated from the genomic sequence) and included five known genes (*MPO*, *MUL*, *RPS6KB1*, *APPBP2*, and *TRAP240*) (Fig. 7). Five of the overexpressed sequences (the *MUL*, *RPS6KB1*, *APPBP2*, and *TRAP240* genes as well as one uncharacterized EST, AA806470) mapped in the common amplified regions identified by FISH in the breast cancer cell lines.

Discussion

Studies by comparative genomic hybridization have indicated that the 17q23 region is one of the most commonly amplified regions in breast cancer and may therefore harbor genes important for breast cancer development and progression. Previous studies by us and others have implicated five known genes (*RPS6KB1*, *RAD51C*, *PAT1/APPBP2*, *SIGMA1B*, and *TBX2*, discussed above in Section II) as putative target genes for this amplification (Bärlund *et al.*, *J. Natl. Cancer Inst.* 92:1252-1259, 2000, Couch *et al.*, *Cancer Res.* 59:1408-1411, 1999; Bärlund *et al.*, *Cancer Res.* 60:5340-5344, 2000; Wu *et al.*, *Cancer Res.* 60:5371-5375, 2000). Presented here are results from a detailed analysis of the 17q23 amplicon in breast cancer cell lines and a large series of primary breast tumors. This study utilized cDNA microarray technology for comprehensive evaluation of copy numbers and expression levels of almost all genes from the 17q23 chromosomal region.

A 4 Mb contig was constructed for the 17q23 region and the precise locations of 17 known genes and 26 transcripts identified from the GeneMap'99 were determined within this contig. To obtain a more complete representation of all genes from this region, the genomic sequence generated by the Human Genome Project was also utilized. Although the sequence-based analysis did not cover the entire 17q23 region, a large number of transcribed sequences were identified that were not present in the GeneMap'99, indicating the importance of such bioinformatics effort in uncovering and localization of novel genes.

The structure of the 17q23 amplicon was evaluated in detail in six breast cancer cell lines using 22 genomic clones from the contig. Amplicon mapping in MCF7 cell line identified two separate regions of amplification flanked by non-amplified areas. BT-474 cell line showed amplification of a single large segment that overlapped both of the regions identified in MCF7 and therefore allowed the definition of two separate common regions of amplification. A recent study by Wu *et al.* (*Cancer Res.* 60:5371-5375, 2000) also identified two regions of amplification at 17q23 in MCF7 and a single region in BT-474, but these regions did not overlap with each other. This discrepancy is most likely due to the fact that the study by Wu *et al.* relied on the GeneMap'99 for mapping information. Unfortunately, the radiation hybrid map does not represent the true order and position of genes in a given chromosomal region and detailed physical mapping studies are necessary for accurate evaluation of amplicon structures. It is expected that the complete human genome sequence will replace the radiation hybrid map in the near future and will therefore make such very

tedious physical mapping efforts unnecessary.

The 17q23 amplification was also studied in a large set of primary breast tumors using the tissue microarray technology. The amplification frequencies of 10 clones from the contig varied from 3.8% to 12.5%. In 74% of the tumors with amplification, five or more clones were co-amplified suggesting that in most primary tumors a large continuous segment at 17q23 is involved in amplification. However, in a few cases smaller regions of involvement were observed but did not allow definition of a single common region of amplification. These results again differ from those obtained by Wu *et al.* (*Cancer Res.* 60:5371-5375, 2000) who studied amplification patterns of four genes from this region and showed co-amplification in only 2 of the 26 amplified tumors. This difference might be explained by differences in the tumor materials analyzed but is most likely due to technical variances in the interpretation of Southern data as compared to the higher precision of copy number evaluation that is possible by FISH analysis.

To fully analyze the copy number and expression patterns of 17q23 specific genes in breast cancer cell lines, we constructed a custom-made cDNA microarray containing 636 cDNA clones including all known genes from chromosome 17 and 156 transcripts from the 17q23 region. CGH microarray analysis indicated a distinct region of amplification at 17q23 and the expression profiling revealed sixteen consistently overexpressed genes in the 17q23 amplified cell lines. Of these, *MUL*, *RPS6KB1*, *APPBP2*, *TRAP240* and an uncharacterized EST, AA806470, were located in the region found to be commonly amplified in these breast cancer cell lines and may therefore contribute to the more aggressive clinical course observed in breast cancer patients with 17q23 amplified tumors (Bärlund *et al.*, *J. Natl. Cancer Inst.* 92:1252-1259, 2000). However, it is possible that the other highly overexpressed genes, although not located in the common region of amplification, also play a significant role in breast cancer pathogenesis.

The cDNA microarray analysis confirmed the previous data from this group and others, suggesting *RPS6KB1* and *APPBP2* as putative targets for the 17q23 amplification (Bärlund *et al.*, *J. Natl. Cancer Inst.* 92:1252-1259, 2000; Couch *et al.*, *Cancer Res.* 59:1408-1411, 1999; Bärlund *et al.*, *Cancer Res.* 60:5340-5344, 2000; Wu *et al.*, *Cancer Res.* 60:5371-5375, 2000). Of the other previously implicated 17q23 genes, *SIGMA1B* was not included in this survey, *RAD51C* was overexpressed in two cell lines, and *TBX2* did not show overexpression possibly due to the difficulty in quantitation of scarce messages by cDNA microarray.

The present cDNA microarray analysis indicated that several other genes are also activated by the 17q23 amplification. The *MUL* gene was recently cloned and showed to encode for a member of the zinc finger protein family with RING-B-box-Coiled- coil domain (Avela *et al.*, *Nat. Genet.* 25:298-301, 2000). Members of this protein family have various functions including regulation of development and cell proliferation (Saurin *et al.*, *Trends Biochem. Sci.* 21:208-214, 1996; Slack & Ruvkun, *Trends Biochem. Sci.* 23:474-475, 1998), thus making *MUL* a very interesting putative target gene for the 17q23 amplification.

TRAP240 belongs to a large multi-subunit complex of thyroid hormone-receptor associated proteins that in a ligand-dependent manner interact with thyroid receptor and activate transcription of genes involved in diverse biological processes (Mangelsdorf & Evans, *Cell* 83:841-850, 1995). The cDNA microarray survey also implicated four uncharacterized ESTs (AA806470, AA635172, stSG39547, and G29449) from the amplified region as putative target genes.

The following non-limiting examples are illustrative of the scope of the present invention.

EXAMPLE 1:

Expression of Polypeptides

With the disclosure herein of linkage between RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 and cancer, the usefulness of expression and purification of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein(s) by standard laboratory techniques is now revealed. Proteins or polypeptides encoded by the antisense strand of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA can likewise be expressed. After expression, purified RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus protein(s) or polypeptide(s) (*e.g.*, fragments) may be used for functional analyses, antibody production, diagnostics, and patient therapy. Furthermore, it can now be appreciated that the DNA sequence of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA and the corresponding antisense strand can be manipulated in studies to understand how expression of these gene and the function of the corresponding products influences neoplasm and development susceptibility to neoplasm. Mutant forms of the human RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene may be isolated based upon information contained herein, and may be studied in order to detect alteration in expression patterns in terms of relative quantities, tissue specificity and functional properties of the corresponding encoded mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein. Partial or full-length cDNA sequences, which encode the subject protein, may be ligated into bacterial expression vectors. Methods for expressing large amounts of protein from a cloned gene introduced into *Escherichia coli* (*E. coli*) may be utilized for the purification, localization and functional analysis of proteins. For example, fusion proteins consisting of amino terminal peptides encoded by a portion of the *E. coli lacZ* or *trpE* gene linked to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 proteins may be used to prepare polyclonal and monoclonal antibodies against these proteins.

Thereafter, these antibodies may be used to purify proteins by immunoaffinity chromatography, in diagnostic assays to quantitate the levels of protein and to localize proteins in tissues and individual cells by immunofluorescence.

Intact native RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein may also be produced in *E. coli* in large amounts for functional studies. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described in Sambrook *et al.* (Sambrook *et al.*, In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989). Such fusion proteins may be made in large amounts, are easy to purify, and can be used to elicit antibody response. Native proteins can be produced in bacteria by placing a strong, regulated promoter and an efficient ribosome-binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy. Suitable methods are presented in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) and are well known in the art. Often, proteins expressed at high levels are found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described by Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989). Vector systems suitable for the expression of *lacZ* fusion genes include the pUR series of vectors (Ruther and Muller-Hill, *EMBO J.* 2:1791, 1983), pEX1-3 (Stanley and Luzio, *EMBO J.* 3:1429, 1984) and pMR100 (Gray *et al.*, *Proc. Natl. Acad. Sci. USA* 79:6598, 1982). Vectors suitable for the production of intact native proteins include pKC30 (Shimatake and Rosenberg, *Nature* 292:128, 1981), pKK177-3 (Amann and Brosius, *Gene* 40:183, 1985) and pET-3 (Studiar and Moffatt, *J. Mol. Biol.* 189:113, 1986). The resultant fusion proteins may be isolated from protein gels, lyophilized, ground into a powder and used as an antigen. The DNA sequence can also be transferred from its existing context to other cloning vehicles, such as other plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke *et al.*, *Science* 236:806-812, 1987). These vectors may then be introduced into a variety of hosts including somatic cells, and simple or complex organisms, such as bacteria, fungi (Timberlake and Marshall, *Science* 244:1313-1317, 1989), invertebrates, plants (Gasser and Fraley, *Science* 244:1293, 1989), and animals (Pursel *et al.*, *Science* 244:1281-1288, 1989), which cell or organisms are rendered transgenic by the introduction of the heterologous RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA.

For expression in mammalian cells, the cDNA sequence may be ligated to heterologous promoters, such as the simian virus (SV) 40 promoter in the pSV2 vector (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981), and introduced into cells, such as monkey COS-1 cells (Gluzman, *Cell* 23:175-182, 1981), to achieve transient or long-term expression. The stable integration of the chimeric gene construct may be maintained in mammalian cells by biochemical

selection, such as neomycin (Southern and Berg, *J. Mol. Appl. Genet.* 1:327-341, 1982) and mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981).

DNA sequences can be manipulated with standard procedures such as restriction enzyme digestion, fill-in with DNA polymerase, deletion by exonuclease, extension by terminal deoxynucleotide transferase, ligation of synthetic or cloned DNA sequences, site-directed sequence-alteration via single-stranded bacteriophage intermediate or with the use of specific oligonucleotides in combination with PCR.

The cDNA sequence (or portions derived from it) or a mini gene (a cDNA with an intron and its own promoter) may be introduced into eukaryotic expression vectors by conventional techniques. These vectors are designed to permit the transcription of the cDNA in eukaryotic cells by providing regulatory sequences that initiate and enhance the transcription of the cDNA and ensure its proper splicing and polyadenylation. Vectors containing the promoter and enhancer regions of the SV40 or long terminal repeat (LTR) of the Rous Sarcoma virus and polyadenylation and splicing signal from SV40 are readily available (Mulligan *et al., Proc. Natl. Acad. Sci. USA* 78:1078-2076, 1981; Gorman *et al., Proc. Natl. Acad. Sci USA* 78:6777-6781, 1982). The level of expression of the cDNA can be manipulated with this type of vector, either by using promoters that have different activities (for example, the baculovirus pAC373 can express cDNAs at high levels in *S. frugiperda* cells (Summers and Smith, In *Genetically Altered Viruses and the Environment*, Fields *et al.* (Eds.) 22:319-328, CSHL Press, Cold Spring Harbor, New York, 1985) or by using vectors that contain promoters amenable to modulation, for example, the glucocorticoid-responsive promoter from the mouse mammary tumor virus (Lee *et al., Nature* 294:228, 1982). The expression of the cDNA can be monitored in the recipient cells 24 to 72 hours after introduction (transient expression).

In addition, some vectors contain selectable markers such as the *gpt* (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981) or *neo* (Southern and Berg, *J. Mol. Appl. Genet.* 1:327-341, 1982) bacterial genes. These selectable markers permit selection of transfected cells that exhibit stable, long-term expression of the vectors (and therefore the cDNA). The vectors can be maintained in the cells as episomal, freely replicating entities by using regulatory elements of viruses such as papilloma (Sarver *et al., Mol. Cell Biol.* 1:486, 1981) or Epstein-Barr (Sugden *et al., Mol. Cell Biol.* 5:410, 1985). Alternatively, one can also produce cell lines that have integrated the vector into genomic DNA. Both of these types of cell lines produce the gene product on a continuous basis. One can also produce cell lines that have amplified the number of copies of the vector (and therefore of the cDNA as well) to create cell lines that can produce high levels of the gene product (Alt *et al., J. Biol. Chem.* 253:1357, 1978).

The transfer of DNA into eukaryotic, in particular human or other mammalian cells, is now a conventional technique. The vectors are introduced into the recipient cells as pure DNA (transfection) by, for example, precipitation with calcium phosphate (Graham and van der Eb, *Virology* 52:456-467, 1973) or strontium phosphate (Brash *et al., Mol. Cell Biol.* 7:2013, 1987),

electroporation (Neumann *et al., EMBO J* 1:841, 1982), lipofection (Felgner *et al., Proc. Natl. Acad. Sci USA* 84:7413, 1987), DEAE dextran (McCuthan *et al., J. Natl. Cancer Inst.* 41:351, 1968), microinjection (Mueller *et al., Cell* 15:579, 1978), protoplast fusion (Schafner, *Proc. Natl. Acad. Sci. USA* 77:2163-2167, 1980), or pellet guns (Klein *et al., Nature* 327:70, 1987). Alternatively, the cDNA, or fragments thereof, can be introduced by infection with virus vectors. Systems are developed that use, for example, retroviruses (Bernstein *et al., Gen. Engr'g* 7:235, 1985), adenoviruses (Ahmad *et al., J. Virol.* 57:267, 1986), or Herpes virus (Spaete *et al., Cell* 30:295, 1982). RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoding sequences can also be delivered to target cells *in vitro* via non-infectious systems, for instance liposomes.

These eukaryotic expression systems can be used for studies of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 encoding nucleic acids and mutant forms of these molecules, the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein and mutant forms of these proteins. Such uses include, for example, the identification of regulatory elements located in the 5' region of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene on genomic clones that can be isolated from human genomic DNA libraries using the information contained in the present disclosure. The eukaryotic expression systems may also be used to study the function of the normal complete protein, specific portions of the protein, or of naturally occurring or artificially produced mutant proteins.

Using the above techniques, the expression vectors containing the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene sequence or cDNA, or fragments or variants or mutants thereof, can be introduced into human cells, mammalian cells from other species or non-mammalian cells as desired. The choice of cell is determined by the purpose of the treatment. For example, monkey COS cells (Gluzman, *Cell* 23:175-182, 1981) that produce high levels of the SV40 T antigen and permit the replication of vectors containing the SV40 origin of replication may be used. Similarly, Chinese hamster ovary (CHO), mouse NIH 3T3 fibroblasts or human fibroblasts or lymphoblasts may be used.

The present disclosure thus encompasses recombinant vectors that comprise all or part of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene or cDNA sequences, or all or part of the antisense strand associated with the RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-related locus, for expression in a suitable host. The RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 DNA is operatively linked in the vector to an expression control sequence in the recombinant DNA molecule so that the corresponding polypeptide can be expressed. The expression control sequence may be selected from the group consisting of sequences that control the expression of genes of prokaryotic or eukaryotic cells and

their viruses and combinations thereof. The expression control sequence may be specifically selected from the group consisting of the *lac* system, the *trp* system, the *tac* system, the *trc* system, major operator and promoter regions of phage lambda, the control region of fd coat protein, the early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus and simian virus, the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, the promoter of the yeast alpha-mating factors and combinations thereof.

The host cell, which may be transfected with the vector of this disclosure, may be selected from the group consisting of *E. coli*, *Pseudomonas*, *Bacillus subtilis*, *Bacillus stearothermophilus* or other bacilli; other bacteria; yeast; fungi; insect; mouse or other animal; or plant hosts; or human tissue cells.

It is appreciated that for mutant or variant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 DNA sequences, similar systems are employed to express and produce the mutant product. In addition, fragments of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein can be expressed essentially as detailed above. Such fragments include individual RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein domains or sub-domains, as well as shorter fragments such as peptides. RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein fragments having therapeutic properties may be expressed in this manner also.

EXAMPLE 2

Suppression of Protein Expression

A reduction of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein expression in a transgenic cell may be obtained by introducing into cells an antisense construct based on a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 encoding sequence, respectively, including the reverse complement of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA coding sequence, the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA or gene sequence or flanking regions thereof. For antisense suppression, a nucleotide sequence from the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus, *e.g.* respectively all or a portion of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA or gene or the reverse complement thereof, is arranged in reverse orientation relative to the promoter sequence in the transformation vector. Where the reverse complement of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL,

TRAP240, AA806470, AA635172, stSG39547, or G29449 sequence is used to suppress expression of the corresponding protein, the sense strand of the corresponding RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene or cDNA is inserted into the antisense construct. Other aspects of the vector may be chosen as discussed above (Example 1).

The introduced sequence need not be the full length human RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA or gene or reverse complement thereof, and need not be exactly homologous to the equivalent sequence found in the cell type to be transformed. Generally, however, where the introduced sequence is of shorter length, a higher degree of homology to the native RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus sequence will be needed for effective antisense suppression. The introduced antisense sequence in the vector may be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. The length of the antisense sequence in the vector advantageously may be greater than 100 nucleotides. For suppression of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, transcription of an antisense construct results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the respective endogenous RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene in the cell.

In some instances, it may be advantageous to use, concurrently or in sequence, antisense constructs from more than one of the set of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 encoding molecules, for instance from both RAD51C and S6K, or any other pair of molecules, or three, four, five and so forth up to all ten of the molecules.

Although the exact mechanism by which antisense RNA molecules interfere with gene expression has not been elucidated, it is believed that antisense RNA molecules bind to the endogenous mRNA molecules and thereby inhibit translation of the endogenous mRNA.

Suppression of endogenous RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression can also be achieved using ribozymes. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. The inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Further, dominant negative mutant forms of the disclosed sequences may be used to block endogenous RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 activity.

Suppression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression can be, for instance, used to treat cell proliferative and other disorders caused by abnormalities in the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus.

EXAMPLE 3:

Production of Protein Specific Binding Agents

Monoclonal or polyclonal antibodies may be produced to normal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein or mutant forms of these proteins, or fragments thereof. Optimally, antibodies raised against these proteins or peptides would specifically detect the protein or peptide with which the antibodies are generated. That is, an antibody generated to the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein or a fragment thereof would recognize and bind the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein and would not substantially recognize or bind to other proteins found in human cells.

The determination that an antibody specifically detects the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein is made by any one of a number of standard immunoassay methods; for instance, the Western blotting technique (Sambrook *et al.*, In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989). To determine that a given antibody preparation (such as one produced in a mouse) specifically detects the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein by Western blotting, total cellular protein is extracted from human cells (for example, lymphocytes) and electrophoresed on a sodium dodecyl sulfate-polyacrylamide gel. The proteins are then transferred to a membrane (for example, nitrocellulose) by Western blotting, and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase. Application of an alkaline phosphatase substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a dense blue compound by immunolocalized alkaline phosphatase. Antibodies that specifically detect the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein will, by this technique, be shown to bind to

affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with either inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appear to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al.* (*J. Clin. Endocrinol. Metab.* 33:988-991, 1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony *et al.* (In *Handbook of Experimental Immunology*, Wier, D. (ed.) chapter 19. Blackwell, 1973). Plateau concentration of antibody is usually in the range of about 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (*Manual of Clinical Immunology*, Ch. 42, 1980).

C. Antibodies Raised against Synthetic Peptides

A third approach to raising antibodies against the RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoded proteins or peptides is to use one or more synthetic peptides synthesized on a commercially available peptide synthesizer based upon the predicted amino acid sequence of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein or peptide.

D. Antibodies Raised by Injection of Protein Encoding Sequence

Antibodies may be raised against proteins and peptides of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 by subcutaneous injection of a DNA vector that expresses the desired protein or peptide, or a fragment thereof, into laboratory animals, such as mice. Delivery of the recombinant vector into the animals may be achieved using a hand-held form of the Biolistic system (Sanford *et al., Particulate Sci. Technol.* 5:27-37, 1987) as described by Tang *et al.* (*Nature* 356:152-154, 1992). Expression vectors suitable for this purpose may include those that express a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoding sequence under the transcriptional control of, for instance, either the human β-actin promoter or the cytomegalovirus (CMV) promoter.

Antibody preparations prepared according to these protocols are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample; or for immunolocalization of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

For administration to human patients, antibodies, *e.g.*, RAD51C- S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific monoclonal antibodies, can be humanized by methods known in the art. Antibodies with a desired binding

specificity can be commercially humanized (Scotgene, Scotland, UK; Oxford Molecular, Palo Alto, CA).

EXAMPLE 4:

Nucleic Acid-Based Diagnosis

A major application of the linkage information presented herein is in the area of genetic testing for neoplasms (*e.g.*, breast or other cancers) or predisposition to neoplasms owing to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 deletion, genomic amplification or mutation. The gene sequence of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genes, including intron-exon boundaries and associated 5' and 3' flanking regions, are also useful in such diagnostic methods. Individuals carrying mutations in one or more of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 loci or genes, or having amplifications or heterozygous or homozygous deletions in or of one or more of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 loci or genes, may be detected at the DNA level with the use of a variety of techniques. For such diagnostic procedures, a biological sample of the subject, which biological sample contains either DNA or RNA derived from the subject, is assayed for a mutated, amplified or deleted

amplification mismatch protection (LAMP); or enzymatic mutation scanning (Taylor and Deeble, *Genet. Anal.* 14:181-186, 1999), followed by direct sequencing of amplicons with putative sequence variations.

B. Detection of Known Mutations:

The detection of specific known DNA mutations may be achieved by methods such as hybridization using allele specific oligonucleotides (ASOs) (Wallace *et al.*, CSHL Symp. Quant. Biol. 51:257-261, 1986), direct DNA sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995, 1988), the use of restriction enzymes (Flavell *et al.*, *Cell* 15:25, 1978; Geever *et al.*, 1981), discrimination on the basis of electrophoretic mobility in gels with denaturing reagent (Myers and Maniatis, Cold Spring Harbor Symp. Quant. Biol. 51:275-284, 1986), RNase protection (Myers *et al.*, *Science* 230:1242, 1985), chemical cleavage (Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* 85:4397-4401, 1985), and the ligase-mediated detection procedure (Landegren *et al.*, *Science* 241:1077, 1988). Oligonucleotides specific to normal or mutant RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 sequences can be chemically synthesized using commercially available machines. These oligonucleotides can then be labeled radioactively with isotopes (such as ^{32}P) or non-radioactively, with tags such as biotin (Ward and Langer *et al.*, *Proc. Natl. Acad. Sci. USA* 78:6633-6657, 1981), and hybridized to individual DNA samples immobilized on membranes or other solid supports by dot-blot or transfer from gels after electrophoresis. These specific sequences are visualized by methods such as autoradiography or fluorometric (Landegren *et al.*, *Science* 242:229-237, 1989) or colorimetric reactions (Gebeyehu *et al.*, *Nucleic Acids Res.* 15:4513-4534, 1987). Using an ASO specific for a normal allele, the absence of hybridization would indicate a mutation in the particular region of the gene, or a deleted RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene. In contrast, if an ASO specific for a mutant allele hybridizes to a clinical sample then that would indicate the presence of a mutation in the region defined by the ASO.

C. Detection of Genomic Amplification or Deletion:

Gene dosage (copy number) can be important in neoplasms; it is therefore advantageous to determine the number of copies of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acids in samples of tissue, *e.g.* breast tissue. It can also be advantageous to determine the copy number of certain portions of the disclosed nucleic acids, for instance about the 3'-terminal half or the 3'-terminal third of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA, or of the 5' or 3' region of the gene. Probes generated from the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 encoding sequence (RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 probes or primers), or the reverse complement of these encoding sequence, can be used to investigate and measure genomic dosage in the q23 region of chromosome 17, and more particularly in the

RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene.

Therefore, it may be advantageous to divide the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus into shorter regions and use only certain regions to probe for genomic amplification. By way of example, the human PAT1/APPBP2 locus, cDNA, ORF, coding sequence and gene sequences may be apportioned into about halves or quarters based on sequence length, and the isolated nucleic acid molecules (*e.g.*, oligonucleotides) may be derived from the first or second halves of the molecules, or any of the four quarters. The human PAT1/APPBP2 gene can be used to illustrate this. The portion of the prototypical human PAT1/APPBP2 cDNA shown in GenBank Accession Number NM_006380 is 2385 nucleotides in length and so may be hypothetically divided into about halves (nucleotides 1-1192 and 1193-2385) or about quarters (nucleotides 1-586, 587-1192, 1193-1779 and 1780-2385). The cDNA also could be divided into smaller regions, *e.g.* about eighths, sixteenths, twentieths, fiftieths and so forth, with similar effect.

Another mode of division is to select the 5' (upstream) and/or 3' downstream region associated with the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 cDNA or gene.

Appropriate techniques for measuring gene dosage are known in the art; see for instance, US Patent No. 5,569,753 ("Cancer Detection Probes") and Pinkel *et al.* (*Nat. Genet.* 20:207-211, 1998) ("High Resolution Analysis of DNA Copy Number Variation using Comparative Genomic Hybridization to Microarrays").

Determination of gene copy number in cells of a patient-derived sample using other techniques is known in the art. For example, RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 amplification in cancer-derived cell lines, as well as in uncultured breast cancer or other cells, can be carried out using bicolor FISH analysis. By way of example, interphase FISH analysis of breast cancer cell lines can be carried out as previously described (Bärlund *et al.*, *Genes Chromo. Cancer* 20:372-376, 1997). The hybridizations can be evaluated using a Zeiss fluorescence microscope. Approximately 20 non-overlapping nuclei with intact morphology based on DAPI counterstain are scored to determine the mean number of hybridization signals for each test and reference probe.

For tissue microarrays, the FISH can be performed as described in Kononen *et al.*, *Nat. Med.* 4:844-847, 1998. Briefly, consecutive sections of the array are deparaffinized, dehydrated in ethanol, denatured at 74° C for 5 minutes in 70% formamide/2 x SSC, and hybridized with test and reference probes. The specimens containing tight clusters of signals or >3-fold increase in the number of test probe as compared to chromosome 17 centromere in at least 10% of the tumor cells may be considered as amplified. Microarrays can be constructed as described in WO9944063A2 and WO9944062A1.

Overexpression of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene can also be detected by measuring the cellular level of RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-; or G29449-specific mRNA. mRNA can be measured using techniques well known in the art, including for instance Northern analysis, RT-PCR and mRNA *in situ* hybridization.

The nucleic acid-based diagnostic methods of this disclosure are predictive of proliferation and metastatic potential in patients suffering from breast carcinomas including lobular and duct carcinomas, and other solid tumors, carcinomas, sarcomas, and cancers including carcinomas of the lung like small cell carcinoma, large cell carcinoma, squamous carcinoma, and adenocarcinoma, mesothelioma of the lung, colorectal adenocarcinoma, stomach carcinoma, prostatic adenocarcinoma, ovarian carcinoma such as serous cystadenocarcinoma and mucinous cystadenocarcinoma, ovarian germ cell tumors, testicular carcinomas, and germ cell tumors, pancreatic adenocarcinoma, biliary adenocarcinoma, heptacellular carcinoma, bladder carcinoma including transitional cell carcinoma, adenocarcinoma, and squamous carcinoma, renal cell adenocarcinoma, endometrial carcinomas including adenocarcinomas and mixed Mullerian tumors (carcinosarcomas), carcinomas of the endocervix, ectocervix, and vagina such as adenocarcinoma and squamous carcinoma, tumors of the skin like squamous cell carcinoma, basal cell carcinoma, melanoma, and skin appendage tumors, esophageal carcinoma, carcinomas of the nasopharynx and oropharynx including squamous carcinoma and adenocarcinomas, salivary gland carcinomas, brain and central nervous system tumors including tumors of glial, neuronal, and meningeal origin, tumors of peripheral nerve, soft tissue sarcomas and sarcomas of bone and cartilage. Cells of these tumors that demonstrate a genomic

G29449 protein that result from, for example, mutations in the promoter regions of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, respectively, or mutations within the coding region of the gene which produced truncated, non-functional or unstable polypeptides, as well as from deletions of a portion of or the entire RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene. Alternatively, genomic duplication at the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus may be detected as an increase in the expression level of one or more RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 proteins. Such an increase in protein expression may also be a result of an up-regulating mutation in the promoter region or other regulatory associated with or coding sequence within the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene. Localization and/or coordinated RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression (temporally or spatially) can also be examined using well-known techniques.

The determination of reduced or increased RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus-associated protein levels (*e.g.*, RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein(s) or peptides expressed from the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus, respectively), in comparison to such expression in a normal cell, would be an alternative or supplemental approach to the direct determination of RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoding sequence deletion, amplification or mutation status by the methods outlined above and equivalents. The availability of antibodies specific to RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein(s) or peptides will facilitate the detection and quantitation of cellular RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein(s) by one of a number of immunoassay methods which are well known in the art and are presented in Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). Methods of constructing such antibodies are discussed above, in Example 3.

Any standard immunoassay format (*e.g.*, ELISA, Western blot, or RIA assay) can be used to measure RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 polypeptide or protein levels; comparison is to the corresponding wild-type (normal) RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 levels, and an increase in RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 polypeptide is indicative of an abnormal biological condition such as neoplasia. Immunohistochemical techniques may also be utilized for RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449

polypeptide or protein detection. For example, a tissue sample may be obtained from a subject, and a section stained for the presence of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 using a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific binding agent, respectively (*e.g.*, anti-RAD51C, anti-S6K, anti-PAT1/APPBP2, anti-TBX2, anti-MUL, anti-TRAP240, anti-AA806470, anti-AA635172, anti-stSG39547, or anti-G29449 protein antibody), and any standard detection system (*e.g.*, one which includes a secondary antibody conjugated to horseradish peroxidase). General guidance regarding such techniques can be found in, *e.g.*, Bancroft and Stevens (*Theory and Practice of Histological Techniques*, Churchill Livingstone, 1982) and Ausubel *et al.* (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

For the purposes of quantitating a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein, a biological sample of the subject, which sample includes cellular proteins, is required. Such a biological sample may be obtained from body cells, such as those present in peripheral blood, urine, saliva, tissue biopsy, amniocentesis samples, surgical specimens and autopsy material, particularly breast cells. Quantitation of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein is achieved by immunoassay and compared to levels of the protein found in healthy cells. A significant (*e.g.*, 30% or greater) reduction in the amount of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the cells of a subject compared to the amount of the corresponding protein found in normal human cells could be taken as an indication that the subject may have deletions or mutations in the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene locus, whereas a significant (*e.g.*, 30% or greater) increase would indicate that a duplication (amplification) may have occurred. Deletion, mutation and/or amplification of or within the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus, and substantial under- or over-expression of one or more RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein(s), may be indicative of neoplasia or a predilection to neoplasia or metastasis, and especially breast cancer.

The protein-based diagnostic methods of this disclosure are predictive of proliferation and metastatic potential in patients suffering from breast carcinomas including lobular and duct carcinomas, and other conditions as described above for nucleic acid-based diagnostic methods (Example 4).

EXAMPLE 6:

Knockout and Overexpression Transgenic Animals

Mutant organisms that under-express or over-express RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449, or a combination of two or more of these proteins, are useful for research. Such mutants allow insight into the physiological and/or pathological role of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in a healthy and/or pathological organism. These mutants are "genetically engineered," meaning that information in the form of nucleotides has been transferred into the mutant's genome at a location, or in a combination, in which it would not normally exist. Nucleotides transferred in this way are said to be "non-native." For example, a non-TBX2 promoter inserted upstream of a native TBX2 gene would be non-native. An extra copy of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene on a plasmid, transformed into a cell, would be non-native.

Mutants may be, for example, produced from mammals, such as mice, that either over-express RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 or under-express RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein, or that do not express RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 at all. Over-expression mutants are made by increasing the number of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genes in the organism, or by introducing a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene into the organism under the control of a constitutive or inducible or viral promoter such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter or the metallothionein promoter. Mutants that under-express RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 may be made by using an inducible or repressible promoter, or by deleting the endogenous RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, or by destroying or limiting the function of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, for instance by disrupting the gene by transposon insertion. As above, in any such organism, it may be beneficial to disrupt expression of more than one of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genes.

Antisense genes may be engineered into the organism, under a constitutive or inducible promoter, to decrease or prevent RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus expression, as discussed above in Example 2.

A gene is "functionally deleted" when genetic engineering has been used to negate or reduce gene expression to negligible levels. When a mutant is referred to in this application as having the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene altered or functionally deleted, this refers to the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene, respectively, and to any ortholog of this gene. When a mutant is referred to as having "more than the normal copy number" of a gene, this means that it has more than the usual number of genes found in the wild-type organism, *e.g.*, in the diploid mouse or human.

A mutant mouse over-expressing RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 may be made by constructing a plasmid having the desired RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene driven by a promoter, such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter. This plasmid may be introduced into mouse oocytes by microinjection. The oocytes are implanted into pseudopregnant females, and the litters are assayed for insertion of the transgene. Multiple strains containing the transgene are then available for study.

WAP is quite specific for mammary gland expression during lactation, and MMTV is expressed in a variety of tissues including mammary gland, salivary gland and lymphoid tissues. Many other promoters might be used to achieve various patterns of expression, *e.g.*, the metallothionein promoter.

An inducible system may be created in which the subject expression construct is driven by a promoter regulated by an agent that can be fed to the mouse, such as tetracycline. Such techniques are well known in the art.

A mutant knockout animal (*e.g.*, mouse) from which a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus gene is deleted can be made by removing coding regions of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene from embryonic stem cells. The methods of creating deletion mutations by using a targeting vector have been described (Thomas and Capecch, *Cell* 51:503-512, 1987).

EXAMPLE 7:

Delivery of Specific Nucleic Acid Sequences

The benefits of delivery of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 sequences to cells are now made evident by the present disclosure. Cells to which one or more of these sequences have been delivered are useful for

a variety of purposes, including for instance research and especially research related to cancer diagnosis and prognosis. In some embodiments, these delivery techniques can be used in gene therapy approaches for combating neoplasia, and particularly breast cancer, in subjects.

Retroviruses have been considered a preferred vector for experiments gene delivery, with a high efficiency of infection and stable integration and expression (Orkin *et al., Prog. Med. Genet.* 7:130-142, 1988). The full-length RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene or cDNA can be cloned into a retroviral vector and driven from either its endogenous promoter or from the retroviral LTR (long terminal repeat). Other viral transfection systems may also be utilized for this type of approach, including adenovirus, adeno-associated virus (AAV) (McLaughlin *et al., J. Virol.* 62:1963-1973, 1988), *Vaccinia* virus (Moss *et al., Annu. Rev. Immunol.* 5:305-324, 1987), Bovine Papilloma virus (Rasmussen *et al., Methods Enzymol.* 139:642-654, 1987) or members of the herpes virus group such as Epstein-Barr virus (Margolskee *et al., Mol. Cell. Biol.* 8:2837-2847, 1988).

Recent developments in gene therapy techniques include the use of RNA-DNA hybrid oligonucleotides, as described by Cole-Strauss, *et al.* (*Science* 273:1386-1389, 1996). This technique may allow for site-specific integration of cloned sequences, thereby permitting accurately targeted gene replacement.

In addition to delivery of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 to cells using viral vectors, it is possible to use non-infectious methods of delivery. For instance, lipidic and liposome-mediated gene delivery has recently been used successfully for transfection with various genes (for reviews, see Templeton and Lasic, *Mol. Biotechnol.* 11:175-180, 1999; Lee and Huang, *Crit. Rev. Ther. Drug Carrier Syst.* 14:173-206; and Cooper, *Semin. Oncol.* 23:172-187, 1996). For instance, cationic liposomes have been analyzed for their ability to transfect monocytic leukemia cells, and shown to be a viable alternative to using viral vectors (de Lima *et al., Mol. Membr. Biol.* 16:103-109, 1999). Such cationic liposomes can also be targeted to specific cells through the inclusion of, for instance, monoclonal antibodies or other appropriate targeting ligands (Kao *et al., Cancer Gene Ther.* 3:250-256, 1996).

To reduce the level of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression, gene therapy can be carried out using antisense or other suppressive constructs, the construction of which is discussed in Example 2.

EXAMPLE 8

Kits

Kits are provided which contain the necessary reagents for determining gene copy number (genomic amplification), such as probes or primers specific for the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 gene, as well as written instructions. The instructions can provide calibration curves or charts to compare with the determined (*e.g.*, experimentally measured) values. Kits are also provided to determine elevated expression of mRNA (*i.e.*, containing probes) or expressed RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein (*i.e.*, containing antibodies or other RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-protein specific binding agents). Also provided are kits to determine elevated activity of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

A. Kits For Detection of Genomic Amplification

The nucleotide sequences disclosed herein, and fragments thereof, can be supplied in the form of a kit for use in detection of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genomic amplification and/or diagnosis of neoplasia. In such a kit, an appropriate amount of one or more of the RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific oligonucleotide primers is provided in one or more containers. The oligonucleotide primers may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the oligonucleotide(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, pairs of primers may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers. With such an arrangement, the sample to be tested for the presence of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genomic amplification can be added to the individual tubes and *in vitro* amplification carried out directly.

The amount of each oligonucleotide primer supplied in the kit can be any appropriate amount, depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, the amount of each oligonucleotide primer provided will likely be an amount sufficient to prime several PCR *in vitro* amplification reactions. Those of ordinary skill in the art know the amount of oligonucleotide primer that is appropriate for use in a single amplification reaction. General guidelines may for instance be found in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990),

Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

A kit may include more than two primers, in order to facilitate the PCR *in vitro* amplification of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 genomic sequences, for instance the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 gene or the 5' or 3' flanking region thereof.

In some embodiments, kits may also include the reagents necessary to carry out PCR *in vitro* amplification reactions, including, for instance, DNA sample preparation reagents, appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and deoxyribonucleotides (dNTPs). Written instructions may also be included.

Kits may in addition include either labeled or unlabeled oligonucleotide probes for use in detection of the *in vitro* amplified RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 sequences. The appropriate sequences for such a probe will be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence the probe is complementary to is amplified during the PCR reaction.

It may also be advantageous to provided in the kit one or more control sequences for use in the PCR reactions. The design of appropriate positive control sequences is well known to one of ordinary skill in the appropriate art.

B. Kits for Detection of mRNA Overexpression

Kits similar to those disclosed above for the detection of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 locus genomic amplification can be used to detect RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-associated mRNA overexpression. Such kits include an appropriate amount of one or more of the oligonucleotide primers for use in reverse transcription PCR reactions, similarly to those provided above, with art-obvious modifications for use with RNA.

In some embodiments, kits for detection of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 mRNA overexpression may also include the reagents necessary to carry out RT-PCR *in vitro* amplification reactions, including, for instance, RNA sample preparation reagents (including *e.g.*, an RNAse inhibitor), appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and deoxyribonucleotides (dNTPs). Written instructions may also be included.

Kits may in addition include either labeled or unlabeled oligonucleotide probes for use in detection of the *in vitro* amplified target sequences. The appropriate sequences for such a probe will

[redacted]

be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence the probe is complementary to is amplified during the PCR reaction.

It may also be advantageous to provided in the kit one or more control sequences for use in the RT-PCR reactions. The design of appropriate positive control sequences is well known to one of ordinary skill in the appropriate art.

Alternatively, kits may be provided with the necessary reagents to carry out quantitative or semi-quantitative Northern analysis of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 mRNA. Such kits include, for instance, at least one RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific oligonucleotide for use as a probe. This oligonucleotide may be labeled in any conventional way, including with a selected radioactive isotope, enzyme substrate, co-factor, ligand, chemiluminescent or fluorescent agent, hapten, or enzyme.

C. Kits For Detection of Protein or Peptide Overexpression

Kits for the detection of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein overexpression are also encompassed herein. Such kits will include at least one target (*e.g.*, RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449) protein specific binding agent (*e.g.*, a polyclonal or monoclonal antibody or antibody fragment) and may include at least one control. The RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein-specific binding agent and control may be contained in separate containers. The kits may also include a means for detecting protein target:agent complexes, for instance the agent may be detectably labeled. If the detectable agent is not labeled, it may be detected by second antibodies or protein A, for example, which may also be provided in some kits in one or more separate containers. Such techniques are well known.

Additional components in some kits include instructions for carrying out the assay. Instructions will allow the tester to determine whether RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression levels are elevated. Reaction vessels and auxiliary reagents such as chromogens, buffers, enzymes, etc. may also be included in the kits.

Having illustrated and described the principles of the use of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 biological molecules, both nucleic acids and proteins, it should be apparent to one skilled in the art that the invention can be modified in arrangement and detail without departing from such principles. The scope of the invention is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

We claim:

1. A method of detecting a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid or an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression, comprising detecting the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid or the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression, wherein the biological condition is a neoplasm.

2. The method of claim 1, wherein the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid or abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression comprises an increased amount of nucleic acid or an RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

3. The method of claim 1, wherein the neoplasm comprises a breast tumor.

4. The method of claim 3, wherein the breast tumor comprises a malignant tumor.

5. The method of claim 3, wherein the breast tumor comprises a ductal, lobular, medullary, mucinous, cribriform, tubular, or papillary carcinoma.

6. The method of claim 1, wherein the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid comprises a nucleic acid amplification.

7. The method of claim 6, wherein the nucleic acid amplification comprises a gene amplification.

8. The method of claim 1, wherein the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression comprises an increased expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in a subject.

9. The method of claim 8, wherein the increased expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in a subject is determined by detecting an increased expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in one or more cells of the subject.

10. The method of claim 1, comprising:

(a) reacting at least one target molecule contained in a clinical sample with a reagent comprising a target-specific binding agent to form a target:agent complex, wherein the target is chosen from the group consisting of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and G29449.

11. The method of claim 10, wherein the target molecule is selected from the group consisting of a target-encoding nucleic acid and a target protein.

12. The method of claim 10, wherein the target-specific binding agent is selected from the group consisting of a target oligonucleotide and a target protein-specific binding agent.

13. The method of claim 10, wherein the sample comprises a neoplastic cell.

14. The method of claim 1, wherein the method is used for presymptomatic screening of an individual for neoplasia.

15. The method of claim 1, further comprising the step of *in vitro* amplifying the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid prior to detecting the abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid.

16. The method of claim 9, wherein the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid is *in vitro* amplified using at least one oligonucleotide primer derived from a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein-encoding sequence.

17. The method of claim 16, wherein the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein encoding sequence is that shown in Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449 respectively.

18. The method of claim 16, wherein the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid is *in vitro* amplified by polymerase chain reaction.

19. The method of claim 8, wherein the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid comprises a sequence selected from the group consisting of:

(a) Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and G29449;

(b) nucleic acid sequences having at least 80% sequence identity with (a); and

(c) fragments of (a) or (b).

20. The method of claim 16, wherein at least one oligonucleotide primer comprises at least 10 contiguous nucleotides from a sequence chosen from the group consisting of:

(a) Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and G29449; and

(b) nucleic acid sequences having at least 80% sequence identity with these sequences.

21. An oligonucleotide primer used in the method of claim 16.

22. A recombinant DNA vector comprising the nucleic acid molecule according to claim 21.

23. A recombinant nucleic acid molecule comprising a promoter sequence operably linked to the nucleic acid sequence according to claim 21.

24. The recombinant nucleic acid molecule according to claim 23 wherein the nucleic acid sequence is in antisense orientation relative to the promoter sequence.

25. A cell transformed with the recombinant nucleic acid molecule according to claim 23.

26. A transgenic non-human animal comprising the recombinant nucleic acid molecule according to claim 24.

27. A kit for detecting an overabundance of RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-encoding nucleic acid using the method of claim 15, comprising at least a pair of oligonucleotide primers each comprising at least 10 contiguous nucleotides of Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449.

28. The method of claim 11, wherein the target molecule is a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 encoding sequence.

29. The method of claim 28, wherein the target:agent complexes are detected by nucleotide hybridization.

30. The method of claim 28, wherein the agent is a labeled nucleotide probe.

31. The method of claim 30, wherein the nucleotide probe has a sequence selected from the group consisting of:

(a) Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and G29449;

(b) nucleic acid sequences having at least 80% sequence identity with the sequences of (a); and

(c) fragments of (a) or (b).

32. An oligonucleotide probe used in the method of claim 28.

33. A kit for detecting an overabundance of target nucleic sequence using the method of claim 28, comprising a RAD51C-, S6K-, PAT1/APPBP2-, TBX2-, MUL-, TRAP240-, AA806470-, AA635172-, stSG39547-, or G29449-specific nucleotide binding agent.

34. The kit of claim 33, wherein the agent is capable of specifically binding to

(a) the nucleic acid sequence shown in Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, and G29449;

(b) nucleic acid sequences that differ (a) by one or more conservative amino acid substitutions;

(c) nucleic acid sequences having at least 80% sequence identity to the sequences specified in (a) or (b); or

(d) antigenic fragments of (a), (b), or (c).

35. The method of claim 11, wherein the target molecule is a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

36. The method of claim 35, wherein the complexes are detected by Western blot assay.

37. The method of claim 35, wherein the complexes are detected by ELISA.

38. The method of claim 35, wherein the target protein has a sequence selected from the group consisting of:

(a) Accession Number AF029669, NM_003161, NM_006380, U28049; AB020705, AF117754, AA806470, AA635172, stSG39547, and G29449;

(b) amino acid sequences having at least 80% sequence identity with the sequences of (a); and

(c) fragments of (a) or (b).

39. The method of claim 35, wherein the target specific binding agent is a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific binding agent.

40. The RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific binding agent of claim 40 wherein the agent is selected from the group consisting of: a RAD51C-specific antibody, a S6K-specific antibody, a PAT1/APPBP2-specific antibody, a TBX2-specific antibody, a MUL -specific antibody, a TRAP240-specific antibody, a AA806470-specific antibody, a AA635172-specific antibody, a stSG39547-specific antibody, or a G29449-specific antibody, and functional fragments thereof.

41. The agent of claim 40, wherein the agent is an antibody.

42. The antibody of claim 41, wherein the antibody is a monoclonal antibody.

43. A kit for detecting an overabundance of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein using the method of claim 35, comprising a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific binding agent.

44. The kit of claim 43, wherein the agent is capable of specifically binding to

(a) the amino acid sequence shown in Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449;

(b) amino acid sequences that differ from those specified in (a) by one or more conservative amino acid substitutions;

(c) amino acid sequences having at least 80% sequence identity to the sequences specified in (a) or (b); or

(d) antigenic fragments of (a), (b), or (c).

45. A method of modifying a level of expression of a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in an subject, comprising expressing in the subject a recombinant genetic construct comprising a promoter operably linked to a nucleic acid molecule, wherein the nucleic acid molecule comprises at least 10 consecutive nucleotides of the sequence shown in Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449, respectively, and expression of the nucleic acid molecule changes expression of the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

46. The method of claim 45 wherein the nucleic acid molecule is in antisense orientation relative to the promoter.

47. A method of inhibiting the expression of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449, comprising:

exposing a cell to at least one oligonucleotide of claim 21.

48. A kit for testing a subject for a biological condition associated with an abnormal target nucleic acid or an abnormal target expression, comprising a target-protein-binding agent, wherein the target is selected from the group consisting of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449.

49. The kit of claim 48, wherein the target is RAD51C.

50. The kit of claim 48, wherein the target is S6K

51. The kit of claim 48, wherein the target is PAT1/APPBP2.

52. The kit of claim 48, wherein the target is TBX2.

53. The kit of claim 48, wherein the target is MUL.

54. The kit of claim 48, wherein the target is TRAP240.

55. The kit of claim 48, wherein the target is AA806470.

56. The kit of claim 48, wherein the target is AA635172.

57. The kit of claim 48, wherein the target is stSG39547.

58. The kit of claim 48, wherein the target is G29449.

59. The kit of claim 48, further comprising a means for detecting binding of the target protein binding agent to a corresponding target polypeptide.

60. The kit of claim 48, wherein the subject is a mammal.

61. The kit of claim 60, wherein the mammal is a human.

62. The kit of claim 60, wherein the mammal is a mouse.

63. The kit of claim 48, wherein the biological condition is a proliferative disease.

64. The kit of claim 48, wherein the biological condition is a neoplasm.

65. The kit of claim 48, wherein the biological condition is cancer.

66. The kit of claim 65, wherein cancer is breast cancer.

67. The kit of claim 65, wherein the target-protein-binding agent is an antibody.

68. The kit of claim 67, wherein the antibody is a polyclonal antibody.

69. The kit of claim 67, wherein the antibody is a monoclonal antibody.

70. A kit for detection of a genetic mutation in a sample of nucleic acid, comprising:

(a) a first container containing an oligonucleotide capable of specifically hybridizing with a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid;

(b) a second container containing a PCR primer that effects amplification of a section of nucleic acid complementary to the oligonucleotide in a PCR process; and

(c) a third container containing a fluorescent labeled nucleic acid probe that selectively hybridizes to the oligonucleotide.

71. The kit of claim 70 wherein the fluorescent labeled nucleic acid probe has a length of between 5 and 500 nucleotides.

72. A kit for detection of a genetic mutation in a sample of nucleic acid, comprising:

(a) a first container containing an oligonucleotide capable of specifically hybridizing with a RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid; and

(b) a second container containing a fluorescent labeled nucleic acid probe that is fully complementary to the oligonucleotide.

73. The kit of claim 72, wherein the fluorescent labeled nucleic acid probe has a length of between 5 and 500 nucleotides.

74. An *in vitro* assay kit for determining whether or not a subject has a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression by detecting an overabundance of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in a sample of tissue and/or body fluids from the subject, comprising:

a container comprising an antibody specific for RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein; and

instructions for using the kit,

the instructions indicating steps for

performing a method to detect the presence of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the sample; and

analyzing data generated by the method, wherein the instructions indicate that overabundance of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470,

AA635172, stSG39547, or G29449 protein in the sample indicates that the individual has the biological condition.

75. The kit of claim 74 further comprising a container that comprises a detectable antibody that binds to the RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein specific antibody.

76. An *in vitro* assay kit for determining whether or not a subject has a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 expression, the kit comprising:

a container comprising an antibody specific for RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein;

a container comprising a negative control sample; and

instructions for using the kit, the instructions indicating steps for

performing a test assay to detect a quantity of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in a test sample of tissue and/or bodily fluid from the subject,

performing a negative control assay to detect a quantity of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the negative control sample; and

comparing data generated by the test assay and negative control assay, wherein the instructions indicate that a quantity of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the test sample greater than the quantity of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the negative control sample indicates that the subject has the biological condition.

77. The kit of claim 76 further comprising a container that comprises a detectable antibody that binds to the antibody specific for RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein.

78. An *in vitro* amplification assay kit for determining whether or not an subject has a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid by detecting the presence of a nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in a sample of tissue and/or body fluids from the subject, the kit comprising:

[redacted]

a first container comprising an *in vitro* amplification primer that specifically amplifies the nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein;

a second container comprising a size marker, the size marker being the expected size of amplified DNA if the nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein is present in the sample; and

instructions for using the kit,

wherein the instructions indicate steps for

performing a method to detect the presence of nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the sample and

analyzing data generated by the method,

wherein the instructions indicate that the presence of increased nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the sample indicates that the subject has the biological condition.

79. An *in vitro* assay kit for determining whether or not an individual has a biological condition associated with an abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 nucleic acid, the kit comprising:

a first container comprising PCR primers that specifically amplify nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein;

a second container comprising a size marker, the size marker being the expected size of *in vitro* amplified nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein;

a third container comprising a negative control sample; and

instructions for using the kit,

wherein the instructions indicate steps for

performing a test assay to detect the presence of nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in a test sample of tissue and/or bodily fluid from the subject,

performing a negative control assay to detect the presence of nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 protein in the negative control sample and

comparing data generated by the test assay and negative control assay,

wherein the instructions indicate that the presence of nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in the test

sample greater than the presence of nucleic acid that encodes RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, or G29449 in the negative control sample indicates that the subject has the biological condition.

80. A method of detecting a biological condition associated with an abnormal 17q23 locus, comprising detecting the abnormal 17q23 locus or abnormal protein expression from the locus, wherein the locus consists of a nucleotide sequence of Accession Number AF029669, NM_003161, NM_006380, U28049, AB020705, AF117754, AA806470, AA635172, stSG39547, or G29449 or complementary sequences, or fragments of such nucleotide sequence.

81. The method of claim 80, wherein the biological condition is a neoplasm.

82. The method of claim 80, wherein the biological condition is a cancer.

83. The method of claim 80, wherein the biological condition is breast cancer.

84. A method of breast cancer diagnosis in a subject, comprising:

obtaining at least one sample from the subject; and

detecting an amplification in at least one target nucleic acid in the sample; wherein the at least one target nucleic acid is selected from the group consisting of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and G29449 nucleic acids,

wherein the amplification is indicative of breast cancer in the subject.

85. The method of claim 84, wherein diagnosis includes prognosis.

86. A method of breast cancer diagnosis in a subject, comprising:

obtaining at least one sample from the subject; and

detecting an increase in expression of at least one target protein in the sample; wherein the at least one target protein is selected from the group consisting of RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and G29449 proteins, and

wherein the increase in expression is indicative of breast cancer in the subject.

87. The method of claim 86, wherein diagnosis includes prognosis.

Figure 1

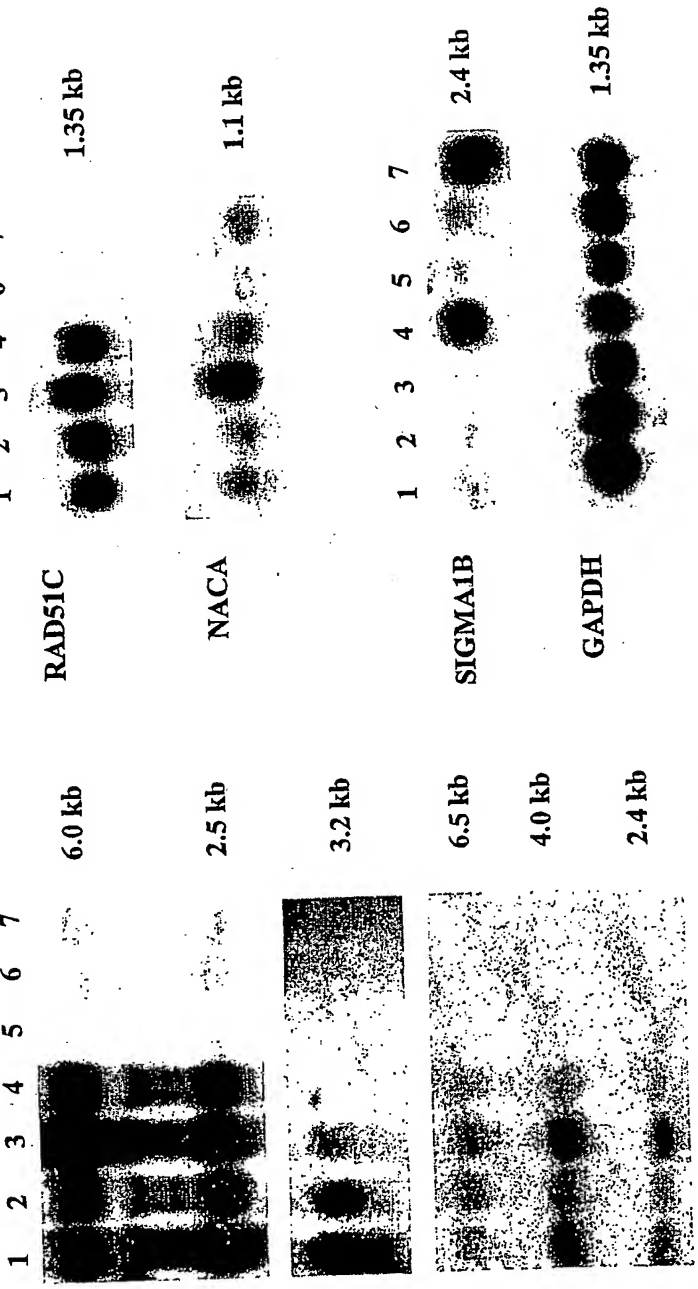

1 2 3 4 5 6 7
S6K
6.0 kb
2.5 kb
TBX2
3.2 kb
PAT1
6.5 kb
4.0 kb
2.4 kb
1 2 3 4 5 6 7
RAD51C
1.35 kb
NACA
1.1 kb
1 2 3 4 5 6 7
SIGMA1B
2.4 kb
GAPDH
1.35 kb

Figure 2

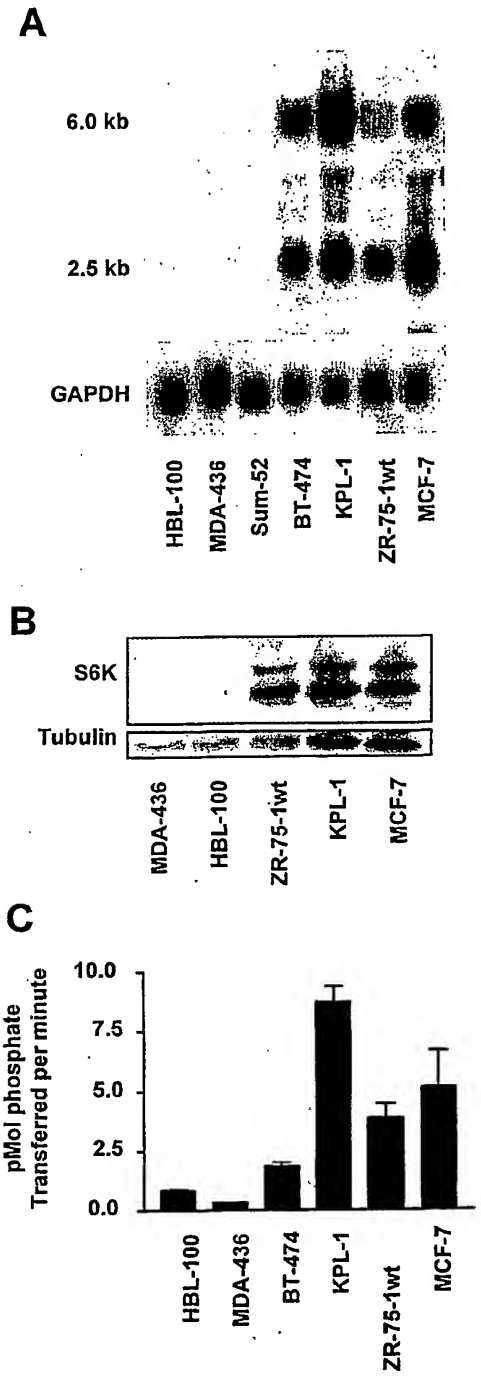

A
6.0 kb
2.5 kb
GAPDH
HBL-100
MDA-436
Sum-52
BT-474
KPL-1
ZR-75-1wt
MCF-7
B
S6K
Tubulin
MDA-436
HBL-100
ZR-75-1wt
KPL-1
MCF-7
C
pMol phosphate Transferred per minute
10.0
7.5
5.0
2.5
0.0
HBL-100
MDA-436
BT-474
KPL-1
ZR-75-1wt
MCF-7

Figure 3

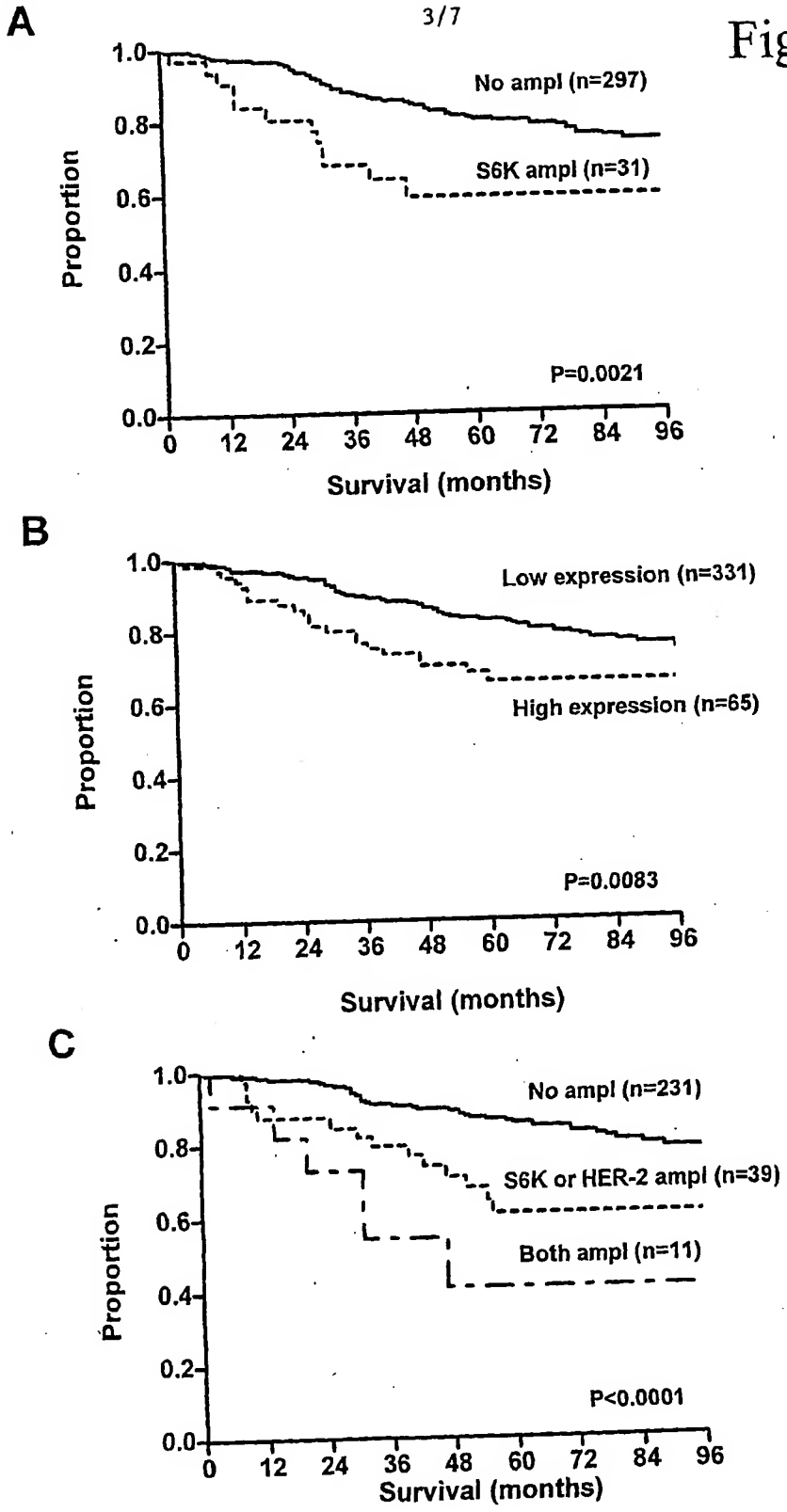

A
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
No ampl (n=297)
S6K ampl (n=31)
P=0.0021
0 12 24 36 48 60 72 84 96
Survival (months)
B
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
Low expression (n=331)
High expression (n=65)
P=0.0083
0 12 24 36 48 60 72 84 96
Survival (months)
C
1.0
0.8
0.6
0.4
0.2
0.0
Proportion
No ampl (n=231)
S6K or HER-2 ampl (n=39)
Both ampl (n=11)
P<0.0001
0 12 24 36 48 60 72 84 96
Survival (months)

Figure 4

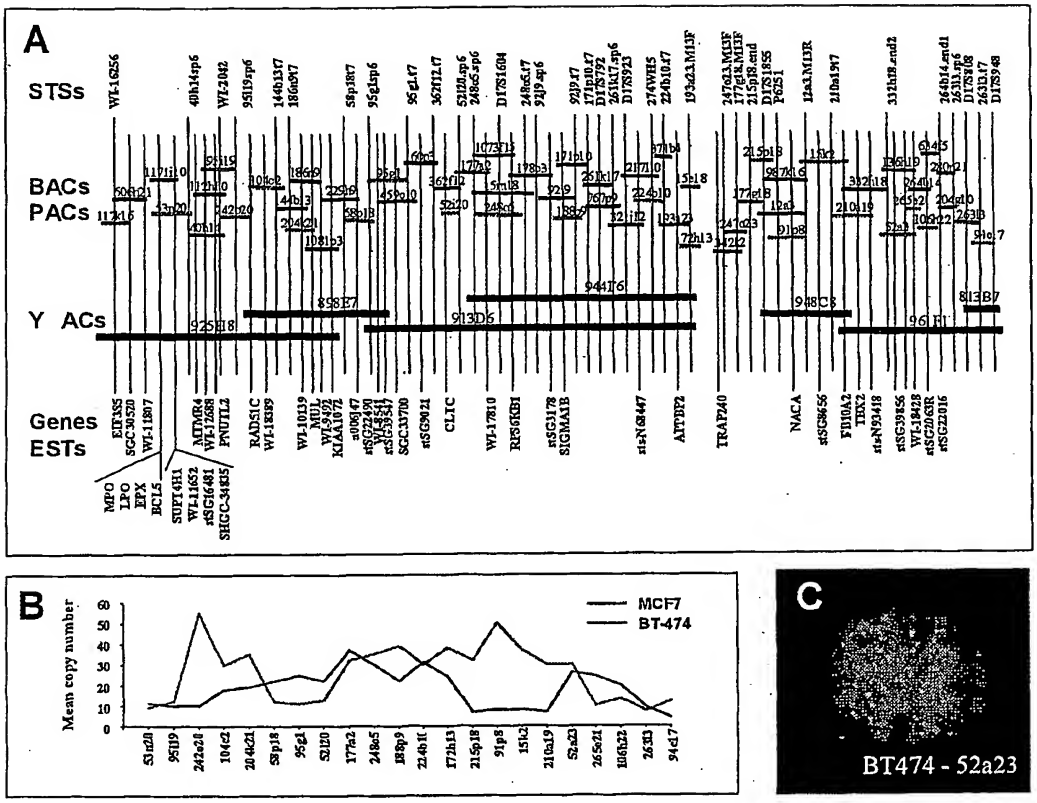

A
STSs
BACs
PACs
Y ACs
Genes
ESTs
B
Mean copy number
MCF7
BT-474
C
BT474 - 52a23

Figure 5

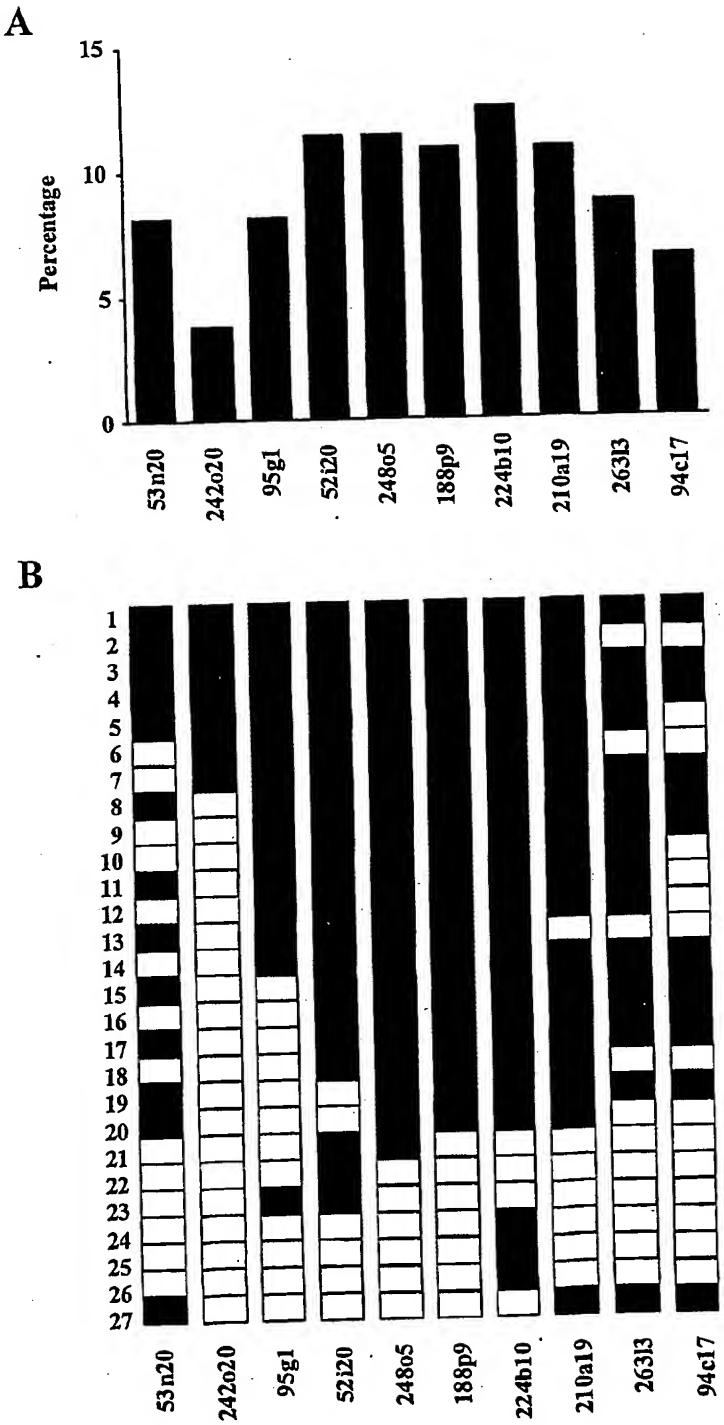

A
Percentage
15
10
5
0
53n20
242o20
95g1
52i20
248o5
188p9
224b10
210a19
263l3
94c17
B
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
53n20
242o20
95g1
52i20
248o5
188p9
224b10
210a19
263l3
94c17

Figure 6

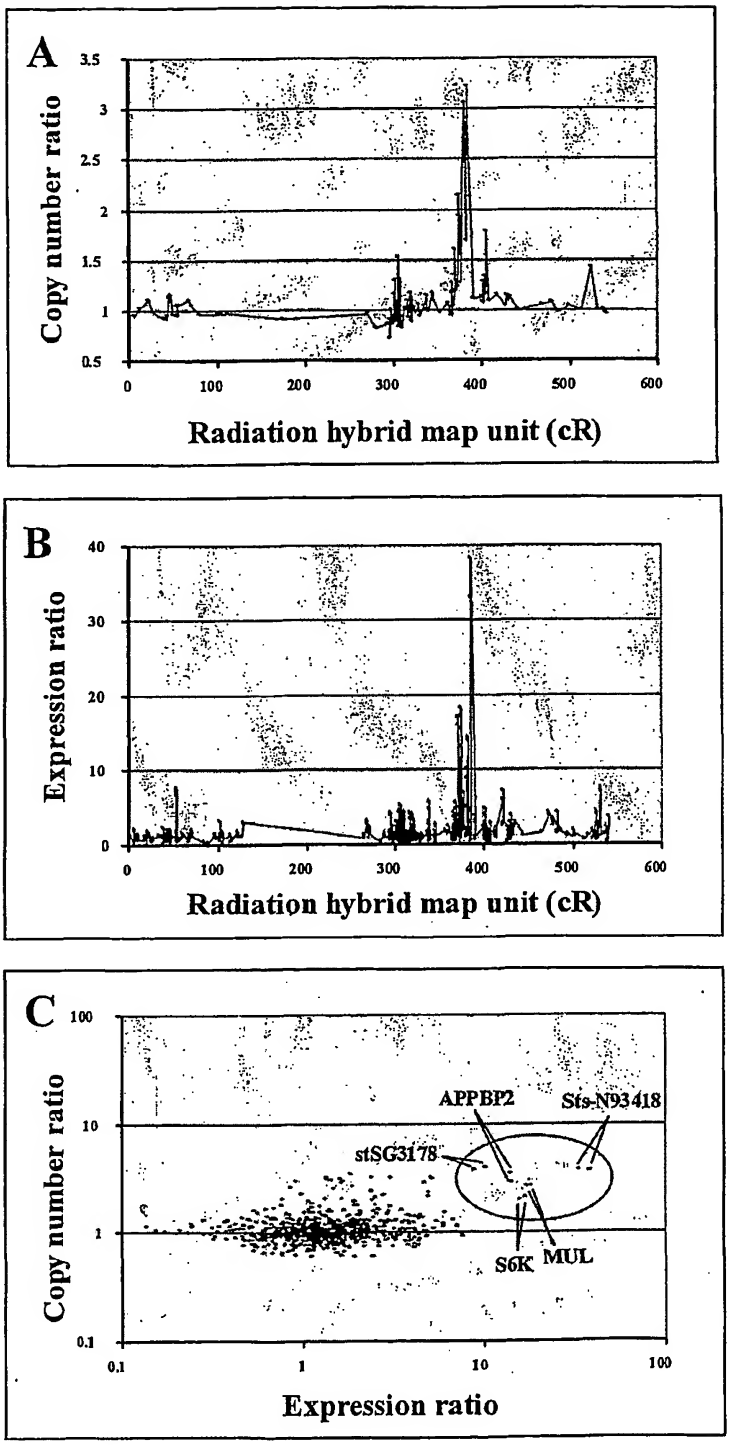
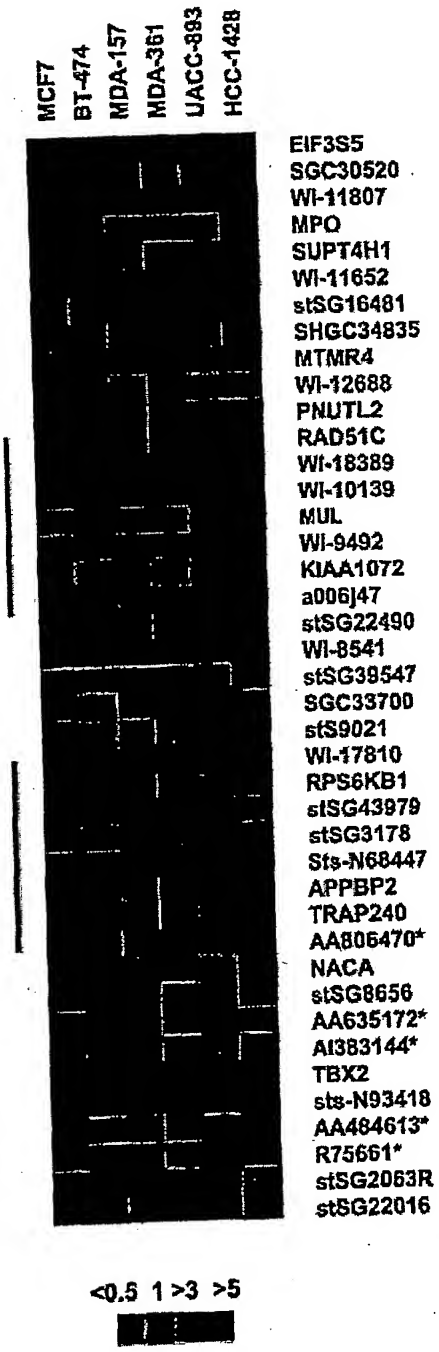

A
Copy number ratio
3.5
3
2.5
2
1.5
1
0.5
0
100
200
300
400
500
600
Radiation hybrid map unit (cR)
B
Expression ratio
40
30
20
10
0
0
100
200
300
400
500
600
Radiation hybrid map unit (cR)
C
Copy number ratio
100
10
1
0.1
APPBP2
Sts-N93418
stSG3178
S6K
MUL
0.1
1
10
100
Expression ratio

Figure 7


MCF7
BT-474
MDA-157
MDA-361
UACC-893
HCC-1428
EIF3S5
SGC30520
WI-11807
MPO
SUPT4H1
WI-11652
stSG16481
SHGC34835
MTMR4
WI-12688
PNUTL2
RAD51C
WI-18389
WI-10139
MUL
WI-9492
KIAA1072
a006J47
stSG22490
WI-8541
stSG39547
SGC33700
stS9021
WI-17810
RPS6KB1
stSG43979
stSG3178
Sts-N68447
APPBP2
TRAP240
AA806470*
NACA
stSG8656
AA635172*
AI383144*
TBX2
sts-N93418
AA484613*
R75661*
stSG2063R
stSG22016
<0.5 1 >3 >5

different in some manner to a normal (wildtype) nucleic acid. Such abnormality includes but is not necessarily limited to: (1) a mutation in the nucleic acid (such as a point mutation or short deletion or duplication of a few to several nucleotides; these may be referred to as polymorphisms); (2) a mutation in the control sequence(s) associated with that nucleic acid such that replication or expression of the nucleic acid is altered (such as the functional inactivation of a promoter; these also may be referred to as polymorphisms); (3) a decrease in the amount or copy number of the nucleic acid in a cell or other biological sample (such as a deletion of the nucleic acid, either through selective gene loss or by the loss of a larger section of a chromosome or under expression of the mRNA); and (4) an increase in the amount or copy number of the nucleic acid in a cell or sample (such as a genomic amplification of part or all of the nucleic acid or the overexpression of an mRNA), each compared to a control or standard. It will be understood that these types of abnormalities can co-exist in the same nucleic acid or in the same cell or sample; for instance, a genomic-amplified nucleic acid sequence may also contain one or more point mutations. In addition, it is understood that an abnormality in a nucleic acid may be associated with, and in fact may cause, an abnormality in expression of the corresponding protein.

Abnormal protein expression, such as abnormal RAD51C, S6K, PAT1/APPBP2, TBX2, MUL, TRAP240, AA806470, AA635172, stSG39547, and/or G29449 protein expression, refers to expression of a protein that is in some manner different to expression of the protein in a normal (wildtype) situation. This includes but is not necessarily limited to: (1) a mutation in the protein such that one or more of the amino acid residues is different (*e.g.*, caused by a polymorphism); (2) a short deletion or addition of one or a few amino acid residues to the sequence of the protein (*e.g.*, caused by a polymorphism); (3) a longer deletion or addition of amino acid residues, such that an entire protein domain or sub-domain is removed or added; (4) expression of an increased amount of the protein, compared to a control or standard amount; (5) expression of an decreased amount of the protein, compared to a control or standard amount; (6) alteration of the subcellular localization or targeting of the protein; (7) alteration of the temporally regulated expression of the protein (such that the protein is expressed when it normally would not be, or alternatively is not expressed when it normally would be); and (8) alteration of the localized (*e.g.*, organ or tissue specific) expression of the protein (such that the protein is not expressed where it would normally be expressed or is expressed where it normally would not be expressed), each compared to a control or standard.

Controls or standards appropriate for comparison to a sample, for the determination of abnormality, include samples believed to be normal as well as laboratory values, even though possibly arbitrarily set, keeping in mind that such values may vary from laboratory to laboratory. Laboratory standards and values may be set based on a known or determined population value and may be supplied in the format of a graph or table that permits easy comparison of measured, experimentally determined values.

THIS PAGE BLANK (USPTO)